2011

THE INTERNATIONAL JOURNAL

cognitive technology®

Dedicated to the enhancement of cognitive functioning through the collaboration of basic and applied research

SELECTED PAPERS FROM THE 10TH BI-ANNUAL INTERNATIONAL CONFERENCE ON NATURALISTIC DECISION MAKING

GUEST EDITORS
MICHELLE HARPER, APTIMA INC.
LEE SCIARINI, NAVAL AIR WARFARE CENTER TRAINING SYSTEMS DIVISION

Volume 16 • Issue 2 • 2011
Special Issue

Published by:
PMI Practical Memory Institute®

Co-Sponsored by:
The Society for Applied Research in Memory and Cognition

ISSN: 1542-3182

ADVERTISEMENT

PMI Practical Memory Institute®
Founded in 1995

Dedicated to helping individual improve memory health and fitness, PMI publishes the Memory Works® memory skill-building CD-ROM series providing individualized learning experiences which are entertaining and educationally effective.

MemoryZine.com is PMI's web site for authoritative information on memory improvement. In addition to providing help for those who what, or need, to enhance their own ability to learn and remember, the web site presents information for caregivers, cognitive rehabilitation clinicians and therapists, and researchers.

The Memory Works® Series
Memory Works® refers to a series of programs that have been developed to help people improve different aspects of their memory performance**.** These programs, a direct result of NIH funded research, redefine memory training. These interactive programs are effective because users engage in experience-based learning and overcome factors that interfere with memory training. Using interactive multimedia technology including animations and 3-D virtual environments, people acquire memory skills at their own pace, in the privacy of their own offices or homes, and with as much practice as they desire. In addition, **Memory Works®** software is also being used by professionals in clinical practice for cognitive rehabilitation of stroke and brain injury victims. **The Memory Works®** products are available at **MemoryZine.com** and include:

- **Best Intentions®** Offers strategies and practice for common memory challenges such as recalling appointments, taking medications, completing errands, attending meetings, remembering chores, and meeting deadlines.
- **Names & Faces®** Instructs users to employ easy-to-learn strategies and practice to better remember the names and faces of individuals.
- **Facts & Figures®** Teaches how to better remember important facts, dates, and long series of numbers by using simple techniques and gaming practices.
- **Nature of Memory®** Explains how the brain learns and remembers, how memory can be affected by outside situations or condition, and how to understand and use external memory devices to improve one's own memory performance.
- **Memory Monitor®** Provides users with a means to examine those factors that determine an individual's readiness for memory challenges. By taking account of a person's lifestyle, fatigue, health, stress, diet, and motivation, the **Memory Monitor®** gauges a person's progress at improving his or her memory fitness over the lifespan.

TO SUBSCRIBE: The International Journal Cognitive Technology (ISSN 1542-3182) is published bi-annually by PMI Practical Memory Institute®. PMI Practical Memory Institute®, Cognitive Technology®, Memory Works®, MemoryZine®, Best Intentions®, and Memory Monitor® are registered trademarks of Compact Disc Incorporated. Download order form or purchase subscription directly using paypal at www.cognitivetechnologyjournal.com.

Single copy process: US $60.00. Yearly subscriptions: US $80.00 individual, $200.00 Institutional. Overseas postage charges: add US $10.00/year. Access to online pdfs for the current year: $25.00 (Available to individuals only). **ORDERS**: (payable to) Cognitive Technology C/O Practical Memory Institute Subscriptions Department, 1610 South Fifth Street, Terre Haute, IN 47802 Entire contents copyright 2010 PMI. Abstracts of articles are published by PMI on the Internet Web site. (www.cognitivetechnoloyjournal.com). No portion of the contents of this periodical publication may be reproduced without written permission of the Editors. (email: brad.brubaker@indstate.edu).

The International Journal ***Cognitive Technology®*** is the official journal of PMI Practical Memory Institute® and is co-sponsored by the Society for Applied Research on Memory and Cognition (SARMAC), and published by Compact Disc Inc. under the auspices of the Psychology Department of Indiana State University, Terre Haute, Indiana.

Cognitive Technology is indexed by Psych INFO of the American Psychological Association.

THE INTERNATIONAL JOURNAL cognitive technology®

EDITOR
Stephen M. Fiore, Ph.D.
University of Central Florida, Orlando, FL

MANAGING EDITOR
Brad Brubaker, Ph.D.
Practical Memory Institute and
Indiana State University, Terre Haute, IN

ASSOCIATE EDITORS

• **MODELING & SIMULATION** •
Rudy Darken, Ph.D.
Naval Postgraduate School, Monterey, CA

• **HUMAN FACTORS** •
Florian Jentsch, Ph.D.
University of Central Florida, Orlando, FL

• **COGNITIVE REHABILITATION** •
Rick Parenté, Ph.D.
Society for Cognitive Rehabilitation and
Towson State University, Towson, MD

• **COGNITIVE SYSTEMS ENGINEERING** •
Jean MacMillan, Ph.D.
Aptima Incorporated, Woburn, MA

• **COGNITIVE MODELS AND AGENT TECHNOLOGIES** •
Gregory Trafton, Ph.D.
Naval Research Laboratory, Washington, DC

PRODUCTION EDITORS

Stephanie Lackey, Ph.D.
Joy M. Martinez
University of Central Florida,
Institute for Simulation and Training, Orlando, FL

PUBLISHER

Robert Rager
Compact Disc Incorporated

EDITORIAL BOARD

Herman Buschke, **M.D.,** Einstein Medical Center
Haydee Cuevas, Ph.D., SA Technologies
John Dunlosky, Ph.D., Kent State University
Frank Durso, Ph.D., Texas Tech University
Jeffrey A. Gibbons, Ph.D., Christopher Newport University
Saul Greenberg, Ph.D., University of Calgary
Michael Gruneberg, **Ph.D.,** University of Wales
J. Preston Harley, **Ph.D.,** Neuropsychology Institute
Douglas Herrmann, Ph.D., Practical Memory Institute
Paula Hertel, Ph.D., Trinity University
Robert Hoffman, Ph.D., Institute for Human and Machine Cognition
Tom Landauer, Ph. D., University of Colorado
H. Chad Lane, Ph.D., University of Southern California, Institute for Creative Technologies
Steve Lewandowsky, **Ph.D.,** University of Western Australia
Elizabeth Loftus, Ph.D., University of California at Irvine
Kit Malia, B.Ed., **M.Phil.,** CPCRT, Defence Services Medical Rehabilitation Centre
Christopher Mayhorn, **Ph.D.,** North Carolina State University
Rudy McDaniel, Ph.D., University of Central Florida
Cathy McEvoy, Ph.D., University of South Florida
Leslie Miller, **Ph.D.,** LanneM Talent Management Consulting
Roger Morrell, **Ph.D.,** Practical Memory Institute
Peter Morris, **Ph.D.,** University of Lancaster
Ray Nickerson, **Ph.D.,** Tufts University
Donald Norman, **Ph.D.,** University of California at San Diego
David Payne, **Ph.D.,** Educational Testing Service
Kathy Pezdek, **Ph.D.,** Claremont University
Dana Plude, **Ph.D.,** Center for Scientific Review
Douglas Raybeck, **Ph.D.,** Hamilton College
George Rebok, **Ph.D.,** The Johns Hopkins University
John T. E. Richardson, **Ph.D.,** The Open University
Albert Rizzo, **Ph.D.,** University of Southern California
Virgil Sheets, Ph.D., Indiana State University
Robert J. Sternberg, **Ph.D.,** Tufts University
Daniel D. Suthers, Ph.D., University of Hawai`i at Manoa
Rodney J. Vogl, Ph.D., Christian Brothers University
Jennifer Wiley, Ph.D., University of Illinois at Chicago

THE INTERNATIONAL JOURNAL
cognitive
technology®

CT-16(2) • VOLUME 16 • ISSUE 2

Special Issue

Selected Papers from the 10th Bi-Annual
International Conference on
Naturalistic Decision Making

Guest Editors

Michelle Harper, Aptima Inc.

Lee Sciarini, Naval Air Warfare Center
Training Systems Division

SELECTED PAPERS

INSTRUCTIONS TO AUTHORS
and SUBSCRIPTION FORMS
can be found at www.cognitivetechnologyjournal.com

From the Editor

Dear Readers,

Welcome to the next issue of *Cognitive Technology*. We are pleased to bring you a special issue based upon selected papers from the 10th Bi-Annual International Conference on Naturalistic Decision Making. For this special issue we have guest editors, Michelle Harper, Aptima Inc. and Lee Sciarini, Naval Air Warfare Center Training Systems Division.

The NDM conferences have long provided an important forum where an eclectic community of scholars can examine how outstanding national and international needs are addressed by recent advances in theory and methods. The 10th International Conference on Naturalistic Decision Making (NDM-2011) continued this tradition and brought together researchers and practitioners from diverse domains who seek to understand and improve how people actually perform cognitively complex functions in demanding situations. The NDM community represents an important interdisciplinary group of researchers united by their study of human performance in situations marked by time pressure, uncertainty, vague goals, high stakes, team and organizational constraints, changing conditions, and varying amounts of experience. As such it continues to be the premier forum for presenting work exploring complex cognition as it occurs in dynamic and real-world contexts.

Cognitive Technology is pleased to bring you this selection of papers from that conference. We continue to help the journal, and the field of applied cognition, take on the responsibility to steward the important interdisciplinary growth in these ever expanding areas of cognitive technology.

Sincerely,

Stephen M. Fiore

Stephen M. Fiore, Ph.D.

Editor, *Cognitive Technology*

Situating Cyber Situation Awareness

Michael D. McNeese
Pennsylvania State University
College of Information Sciences and Technology

Nancy J. Cooke
Arizona State University
College of Technology and Innovation

Michael A. Champion
Arizona State University
College of Technology and Innovation

We present a framework for understanding cyber situation awareness that is ecologically inspired. The view that situation awareness involves interactions between the physical, psychological, and environmental realms reflects a compromise between perspectives, which locate situation awareness primarily in the machine or in the human. Implications of this view are discussed.

KEYWORDS: Situation Awareness, Cyber Situation Awareness, Cyber Security, Cyber Security Analysts

WHAT IS CYBER SECURITY?

Cyber security occurs within a sociotechnical system of vast distributed arrays of computers, servers, and analysts. Its essence cannot be captured from a technology-only approach, but it needs to be examined more broadly in a way that incorporates cognitive, organizational, environmental-contextual and technological demands in an interdisciplinary manner. Our consideration necessarily implies that cyber-security is both defined and acted upon by humans for humans, through the use of computer-based tools. Compromised systems can result in a loss of human, informational, and/or material resources either directly or indirectly. Humans, working conjunctively with information availability and technological advancement, have the possibility of improving performance in cyber security, if cyber-based technology is designed with the human in mind from the beginning. Certainly our research approaches (e.g., Living Laboratory Framework, McNeese, 1996; synthetic task environments, Cooke & Shope, 2004; communication analysis, Cooke & Gorman, 2009, and other socio-technical approaches, Rasmussen, Pejtersen, & Goodstein, 1994; Rouse, 1988) have utilized multiple methods of analysis and synthesis to improve system designs from multiple perspectives, for a variety of work domains. The focus of this paper is to understand Situation Awareness (SA) as it relates to cyber security. More specifically, this paper focuses on the often-conflicting views regarding the locus of situation awareness.

Cognitive Systems Engineering and Cyber Security

There are a number of cognitive systems engineering studies that have explored the domain of cyber security using cognitive task analysis and other similar analytic techniques in order to better understand the cognitive challenges associated with cyber analysis both at the individual and team levels. Branlot, Morison, Finco, Gertman, Le Blanc, and Woods (2011) found that the cyber security task is indeed cognitively challenging in that it involves data overload, competing goals, and requires iterative discovery of knowledge. This is compounded by the fact that the representation of emerging knowledge (owing to the virtual world component of cyber security) is often stale. Other studies (e.g., Brown, 1998; Kraemer, Carayon, & Duggan, 2004) also reveal the challenges associated with teamwork in the cyber domain that include socio-psychological challenges, communication complexities and adversarial team-to-team work that lies at the foundation of many cyber security attacks.

Certainly, SA is key in such a dynamic and uncertain domain. Much work in cognitive systems engineering and computer science has been dedicated to developing tools, algorithms, and visualizations to improve cyber SA (e.g., Mahoney, Roth, Steinke, Pfaulz, Wu, & Farry, 2011). Effective interventions however are predicated on a firm understanding of situation awareness in the cyber domain.

What is Situation Awareness?

One very relevant publication describing different theories and methods of SA was the 1995 special issue of *Human Factors*, Volume 37. There are various definitions of SA and most of the early works emanate from the different periods within the 1980s (e.g., Endsley, 1988; Fracker, 1988; Rouse & Rouse, 1983; Weiner & Curry, 1980). In fact, the concept evolved from earlier work in human factors, aviation, and cognition. The first two authors of this paper have both worked in various aspects of SA over the last 25 years in differing capacities. The first author's first publication involving automation and SA appeared in the Human Factors and Ergonomics Society's annual meetings proceedings paper in 1985 (McNeese, Warren, & Woodson, 1985). The seminal work in SA by Mica Endsley states that SA "is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1995, p. 36). Indeed, this definition, the framework associated with it, and the measurement methodology utilized are the standard bearers for understanding the concept and what it means in many different applications areas. A recent review by Wickens (2008) provides an extensive analysis of the pros and cons associated with the SA construct.

What is Cyber Situation Awareness?

What does SA mean in the cyber security domain? In a practical sense this term refers to actionable understanding of the cyber threat situation at any one point in time and consistent with Endsley's (1995) definition this involves cognitive processes of perception, comprehension, and projection. D'Amico, Whitley, Tesone, O'Brien, and Roth (2005) emphasize that maintaining actionable understanding in the cyber domain requires information fusion, specific knowledge building, maintaining and tracking everything, and the coordination of multiple mental models. They suggest that these processes can be facilitated through the use of role-based visualization.

The need for SA in the cyber domain and the challenges associated with it are clear, but what is less clear is where the SA resides. The locus of SA prescribes specific types of interventions to improve it. As cognitive engineers and human factors professionals we are focused on the person as the locus of SA.

Unfortunately, the locus of SA may be situated differently by those who do not share this bias. In particular, in the cyber domain, rife with algorithms, hardware, and software tools, SA is often depicted as being within the computer system. It may reside in the data that serve as input to the analyst and the objective becomes collecting as much data as possible about the situation in order to maximize SA (Jajodia & Noel, 2010). Although there is some truth that data or observations of the situation are inextricably tied to the understanding of the situation, it is, however, not guaranteed that increasing such data would automatically improve SA. In actuality, it could be that increasing data obfuscates our understanding of the situation. One of the associated myths that arise is that more data equals a better opportunity to solve cyber-security problems. Although this may appear to be true on the surface level, more data typically increases the complexity of situations, which results in the "interpretation dilemma" for humans. That is, understanding of complexity often requires sufficient time to interpret and fuse data to achieve information and in turn, SA. However the effective time available in real world events may be very short. This situation is analogous to finding several differently colored needles in a haystack, in which all contain a segment of value or truth, but together form the entire basis of value or truth. Obviously the additional element of uncertainty weighs heavily in complex situations involving time pressure. Further complicating this dilemma is that porous nature of data uptake (i.e., the feedback loop on any strand of data may not be all that recent and may be challenged in terms of whether the data are coming from a "trusted source.")

In some cases SA is likened to a display or visualization. For example, the objective may be to provide the analysts with "situation awareness displays" or common operating pictures (Ackerman, 2010). This too is unsatisfactory because we cannot guarantee that a display will provide adequate SA. In fact, similar to the data-as-situation-awareness argument, presenting all of the information that we can about a situation on a display may not help at all.

In the cyber domain there are algorithms and tools for filtering and fusing data and providing alerts regarding potential threats. These are sometimes referred to as "situation awareness algorithms" (Barford, Dacier, Dietterich et al., 2010). These algorithms or automations are intended to increase SA. However, there is no guarantee that the automation facilitates SA. In fact, it has been documented that increasing automation too much can have a trade-off within SA (e.g., Endsley, 1995; Endsley and Kiris, 1995; Wikens, 2002).

As cognitive scientists we see a problem. The human is missing from the SA equation (or at least slighted) in these examples. Some may argue that SA needs to fully reside in the human. The fixing of SA in the human is equally suspect. Just as the cockpit knows its speed (Hutchins, 1995) the computer can have SA, but this only makes sense when the human and machine come together (McNeese, 1986). The viewpoint of SA that is heavily influenced by cognitive psychology (e.g., the main tenets of Endsley's theory of SA (1995)—perception, comprehension, and projection—are traditional elements of cognitive psychological processes, or products of cognition) does not give adequate credence to the role of context and ecological constraints that mutually specify behavior and adaptivity. Therefore, the alternative approach that we take emphasizes the interconnectivity and relationships that can proceed from physical to psychological to environmental realms of SA. We propose that the SA, here as exemplified in the cyber domain, is in the interaction between human and machine.

Situating Situation Awareness

At a conceptual level it is best to think of SA as a sphere that surrounds people as they think, act, and move within their world of being. They may *think* about a given situation that is impending upon them and all the aspects surrounding a given situation, *act* in accordance with what they are trying to achieve at both the moment and what might come next, and *move* in concert with the interactions of other humans (both interdependently and independently based on circumstances they find themselves encountering) given what they are facing. Movement can also connote the idea of distributed nature of events, as well as the idea of information flow across events and people. This may be especially relevant for cyber-security given the presence of adversarial attacks and counter measures. To complicate movement even further these attacks may be hidden or disguised wherein spoofing, deception, or fraud is used to disguise movement and identity from human perception. If one considers SA as a "state of readiness" to be directed towards solving problems in the naturalistic world then it is not static, but dynamic. Developing and maintaining SA is a process which can be ephemeral and entail understanding and keeping track of multiple, interwoven events that contain many influences. As events change spatially, temporally, and socially, the complexity of SA can be exceedingly excruciating to maintain and act upon. When we consider "what SA is" and "what it does" we suggest that it is the state of readiness that facilitates an understanding of what the future holds in terms of accessing what you have come to know through the past, to adapt to the currency of the moment in order to accomplish intentionality.

The development and maintenance of SA can exist at a) physical, b) psychological, and c) environmental realms albeit in different orchestrations. The physical realm of awareness is heavily ensconced in sensory-neurological substrates often regulated by attention and bio-physiological levels. At the most basic-level humans are connected to the world first through their sensory apparatus (the sense surround that composes the sphere of awareness). There is some evidence that physical states (e.g., sleep deprivation, fatigue) impact decision-making elements that are related to SA (Harrison, Horne, & James, 2000). Although the exact effects of physical states are typically unknown or complex to specify, it is clear that the physical state is one source that impacts awareness. The physical realm may also reveal specific individual differences that can influence SA (e.g., impulsivity, reading speed). Therein, the physical realm is the most basic in this process of understanding how SA comes to pass.

The psychological realm typically focuses on theoretical models of cognitive and perceptual understanding that exist within the human mind. In many ways, conceptualizing SA requires a multitude of cognitive processes working together to attune the human towards thought, action, and movement into the future. For example, a person may formulate a hypothesis about an emerging situation wherein they could employ cognitive powers such as pattern recognition, learning, memory, reasoning, and judgment to uncover "what is going on" and "what can I do about it".

Upon first consideration, one may consider SA to be heavily analytical or rational. But many situations reveal the use of experiential or intuitive knowledge to also be valuable as part of SA. For example, the work of Kenneth Hammond and associates reveals that humans use analytical and intuitive cognition in solving real world problems (Hammond, Hamm, Grassia, & Person, 1987). Recent research findings also suggest that cognition in general is coupled with emotional states (e.g., frustration, anxiety, fear, and boredom) as well and therein, emotion and mood can influence and color a person's state of readiness in a situation (Hudlicka, 2003; Pfaff & McNeese, 2010).

Finally, the environmental realm points to the role of the context and how situations are grounded in experience as understood by the mind. There is SA in the environment and in this example, in the cyber-security environment. The environmental realm also includes physical nature, the built world, and the social surround that a person's awareness develops in. These realms provide different lenses through which SA may be understood and examined, and often underlie how researchers formulate specific approaches to SA. A comprehensive approach to SA would integrate these realms together to holistically understand what awareness within situations actually means. In spite of these realms being joint possibilities for comprehensively formulating SA, the realm that has produced the mainline view of SA is the psychological realm.

In particular, the approach we take expresses the interdependence and symbiotic relationship between the agent's neurocognitive states (e.g., intentions) and what the environment offers (affordances) to create agent-environment transactions (McNeese, 2001, refer to the work of Gibson, 1986). As an agent works within a constrained environment, patterns and other forms of information are picked up to help specify what needs to be accomplished to achieve an intention. Because actions taken (referred to as effectivities) are made through the imprints of past learning and current pickups, an agent is situationally aware when their actions show signs of adaptivity, and accomplishment of an emerging set of intentions. This is an alternative viewpoint to the traditional SA literature. This definition puts more of an onus on the principle that action is contextually situated and constantly emerging in a dynamic state rather than being a more static model that exists conceptually somewhere "within". It also implies that "situation awareness" is really connected first with the physical being.

Within this framework we can also specify what it means for a team to be situationally aware. Unlike the popular view (Bolstad & Endsley, 2003) that team members who are situationally aware are "on the same page" or sharing the same understanding of the situation, we view team SA as an adaptive coordination among teammates who may each be aware of different aspects of the situation that results in actionable knowledge. This perspective on team SA is described in Gorman, Cooke, and Winner (2006).

Implications

We have utilized an ecological approach that is close conceptually to *situated cognition* frameworks (Young & McNeese, 1995) and as such represents several integrative dispositions:

1) Cognitive processes are not simply easy conceptual tools inside the head but are an agent's neurocognitive states that evolve within real world naturalistic environments that have various constraints and impedance,
2) Naturalistic decision making within these environments is absolutely dynamic and involves adapting to levels of complexity that change dependent on conditions that are operative,
3) There is a natural inclination in human-centered systems to move from the individual agent to multi-agent systems (teams) and that these systems network together into an entwined system of systems that produce distinct qualities that make work complex, yet still highly flexible,
4) When context is stable and routine it is possible for agent(s) to evolve to a point wherein agent-environment transactions are viable, producing a high probability of intentionality for given contexts, that allows the system to be responsive with perturbations to an extent, hence producing levels of automaticity and generalization that make actions highly situated and apropos,
5) When context is messy, unstable, or surprises exist, then the agent seeks and searches for new terrain that enables sense making and evolutionary actions to adapt.

Applying this more ecologically-dominant view of SA emphasizes work contexts that have some of the following characterizations:

- Ill-defined problems, increasing time pressure and impending multiple deadlines,
- The ecology of the environment suggests various goals change dynamically,
- Personnel may hand off work to other shift workers and dynamically reallocate work as needed,
- Information can be perceived directly but information seeking is prevalent within and across teams,
- Physiological determinants can alter attention and therein change action situatedness, rapidly
- Group work is supported by distributed, technological artifacts that need to reflect information flow and movement,
- Information sharing and teamwork interdependencies are necessary for success and to promote collective group agency.

Taken together, this view of SA represents a compromise between the ideas that the locus of SA is in the machine or data (Jajodia & Noel, 2010) versus in the human (Endsley, 1995). Cyber SA requires not only a focus on both machine and human SA, but the intersection of both. As a consequence, algorithms, tools or visualizations developed without an adequate connection to the human user will likely fail to achieve their objectives. Likewise humans cannot achieve objectives of cyber security without the aid of the automation. This compromise position balances the focus of cyber SA between the human and the automation and specifically includes the human-system interactions. As a consequence different metrics for assessing cyber SA in terms of human-automation interaction are suggested and algorithms, tools, and visualizations that exploit this interaction can be explored.

AUTHOR NOTES

Correspondence regarding this paper should be directed to Michael McNeese [MMcNeese@ist.psu.edu].

REFERENCES

Ackerman, R. K. (2010). Network situational awareness looms large in cyberspace. *Signal Online.*

Barford, P., Dacier, M., Dietterich, T. G., Fredrickson, M., Giffin, J., Jajodia, S., … Yen, J. (2010). Cyber SA: Situational Awareness for Cyber Defense. In S. Jajodia, P. Liu, P., Swarup, V., Wang, C. (Eds.), *Cyber Situational Awareness*, *46,* 3-15. New York, NY: Springer**.**

Bolstad, C. A., & Endsley, M. R. (2003). Measuring shared and team situation awareness in the Army's future objective force. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 369–373). Santa Monica, CA: Human Factors and Ergonomics Society.

Branlot, M., Morison, A. M., Finco, G. J., Gertman, D. I., Le Blanc, K. L., & Woods, D. D. (2011). A study of adversarial interplay in a cyber security event. *Proceedings of the 10th International Conference on Naturalistic Decision Making* (NDM 2011), Orlando FL.

Brown, C. E. (1998). Collaborative Systems Technology Assessment for the Network Command Center, presentation given to the Department Head of the 88th Communications Group, Wright-Patterson AFB, OH, March 18, 1998.

Cooke, N. J., & Gorman, J. C. (2009). Interaction-Based Measures of Cognitive Systems. *Journal of Cognitive Engineering and Decision Making, 3,* 27-46.

Cooke, N. J., & Shope, S. M. (2004), Designing a synthetic task environment. In S. G. Schiflett, L. R. Elliott, E. Salas, & M. D. Coovert, *Scaled Worlds: Development, Validation, and Application,* (pp. 263-278). Surrey, England: Ashgate.

D'Amico, A., Whitley, K., Tesone, D., O'Brien, B., & Roth, E. (2005). Achieving cyber defense situational awareness: A cognitive task analysis of information assurance analysts. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting.* (pp.229-233). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 97–101). Santa Monica, CA: Human Factors Society.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37,* 32–64.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37,* 381–394.

Fracker, M. L. (1988). A theory of situation assessment: Implications for measuring situation awareness. *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 102–106). Santa Monica, CA: Human Factors Society.

Gibson, J. J. (1986) *The ecological approach to visual perception*. Hillsdale, NJ: Erlbaum.

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control systems. *Ergonomics, 49,* 1312-1325.

Harrison, Y., & Horne, J. A. (2000). The impact of sleep deprivation on decision making: A review. *Journal of Experimental Psychology: Applied*, *6*(3), 236-249.

Hammond, K. R. Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgement. *IEEE Transactions on Systems, Man and Cybernetics, 17*, 753-767.

Hudlicka, E. (2003). To feel or not to feel: the role of affect in human-computer interaction. *International Journal of Human–Computer Studies*, *59*, 1–32. Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science, 19,* 265-288. doi:10.1016/0364-0213(95)90020-9.

Jajodia, S., & Noel, S. (2010). Topological Vulnerability Analysis. In S. Jajodia, P. Liu, P., Swarup, V., Wang, C. (Eds.), *Cyber Situational Awareness* (pp. 139-154). New York, NY: Springer**.**

Kraemer, S., Carayon, P., & Duggan, R. (2004). Red team performance for improved computer security. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting.* (pp. 1605-1609). Santa Monica, CA: Human Factors and Ergonomics Society.

Mahoney, S., Roth, E., Steinke, K., Pfautz, J., Wu, C., & Farry, M. (2011). A cognitive task analysis for cyber situation awareness. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting.* (pp. 279-283). Santa Monica, CA: Human Factors and Ergonomics Society.

McNeese, M. D. (1986). Humane intelligence: A human factors perspective for developing intelligent cockpits. *IEEE Aerospace and Electronic Systems*, *1* (9), 6-12.

McNeese, M. D. (1996). An ecological perspective applied to multi-operator systems *Human factors in organizational design and management*, *6*, 365-370. The Netherlands: Elsevier.

McNeese, M. D. (2001). The use of ubiquitous computing/computational neuroscience for distributed battlefield management. *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 473-477)*,* Santa Monica, CA: Human Factors and Ergonomics Society.

McNeese, M. D., Warren, R. & Woodson, B. (1985). Cockpit automation technology—A further look. *Proceedings of the 29th Annual Meeting of the Human Factors Society* (pp. 884-888), Santa Monica, CA: Human Factors Society.

Pfaff, M. S., & McNeese, M. D. (2010). Effects of mood and stress on team cognition in a simulated task environment. *Theoretical Issues in Ergonomic Science, 11*(4), 321-339.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). *Cognitive engineering: Concepts and applications*, New York: Wiley.

Rouse, W. B. (1988). Adaptive aiding for human/computer control. *Human Factors*, *30*, 431-438.

Rouse, W. B., & Rouse, S. H. (1983). Analysis and classification of human error. *IEEE Transactions. Systems, Man & Cybernetics*, *13*(4), 539-549.

Weiner, E., & Curry, R. (1980). Flight deck automation: Promises and problems. *Ergonomics, 23,* 995–1012.

Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current Directions in Psychological Science, 11,* 128–133.

Wickens, C. D. (2008). Situation awareness: Review of Mica Endsley's 1995 articles on situation awareness theory and measurement, *Human Factors*, *50*, 397-403.

Young, M. F., & McNeese, M. D. (1995). A situated cognition approach to problem solving. In P. Hancock, J. Flach, J. Caird, & K. Vincente (Eds.), *Local applications of the ecological approach to human-machine systems* (pp. 359-391). Hillsdale, NJ: Erlbaum.

Providing an Option Awareness Basis for Naturalistic Decision Making

Gary L. Klein
The MITRE Corporation
McLean, VA

Jill L. Drury
The MITRE Corporation
Bedford, MA

Mark S. Pfaff
Indiana University
School of Informatics

We describe how we have extended the basis for naturalistic decision making beyond traditional definitions of situation awareness. We have done this by providing a computer-generated decision space that displays the distributions of plausible outcomes for available options. This enables fast visual comparisons producing *option awareness* that augments the mental simulation of recognition primed decision making in complex, uncertain settings: empirically yielding faster, more confident, more robust decisions.

KEYWORDS: Option Awareness, Situation Awareness, Naturalistic Decision Making, Robust Decision Making, Exploratory Modeling.

INTRODUCTION

Situation Awareness (Endsley, 1988) and Naturalistic Decision Making (NDM; G. A. Klein, 1998) have been firmly linked. Endsley writes, "Situation awareness provides the primary input to the decision process and the basis [for] decision strategy selection" (Endsley, 1997, p. 281). Level 1 situation awareness (Endsley, 1995) involves decision makers perceiving information about the environment. Comprehending the meaning of this information is Level 2. Decision makers who have attained Level 3 situation awareness are able to project the state of the environment into the near future. Further, decision makers who have Level 3 situation awareness can use this capability, albeit limited by normal human cognitive capacity, to project into the future the likely outcomes that may occur when choosing one course of action versus another.

However, our research has shown that the cognitively limited consideration of options is often inadequate under situations of deep uncertainty (Lempert, Popper et al. 2003) or when there are many viable options to consider. In fact, our most recent research shows that decision makers, blinded to the actual complexity of a situation, will make overly simple non-robust decisions with high confidence. Furthermore, our work has shown that by providing decision makers with a computer supported visualization of their landscape of options and projected outcomes, decision makers can apply swift perceptual processing to yield faster, more confident, and more robust decisions (Drury et al. 2009a).

Using Hall et al.'s (2007) definitions, we say that situation awareness is supported by the *situation space*: information consisting of facts about the situation. This new visualization of the landscape of options provides information describing the decision options and their desirability relative to one another and therein fulfills Hall's definition of a *decision space*. The extension of perception, comprehension, and projection to this decision space, we have called Option Awareness (Drury et al. 2009a; Klein et al. 2010).

Our research showed that displaying decision spaces does, indeed, provide option awareness (Pfaff et al., 2010b, Drury et al. 2009a). We have demonstrated (Pfaff et al., 2010b) that under circumstances of deep uncertainty these decision space visualizations not only enabled decision makers more often to identify robust options, but to make decisions faster and with more confidence than unaided decision makers. Our results are consistent with those of Nadav-Greenberg and Joslyn (2009), who state that decision makers in naturalistic settings make better decisions when they have uncertainty information as opposed to when the information takes the form of a deterministic forecast.

While others have reported research on visualizing uncertainty in decision making, in general these efforts have either focused on the situation space, or have less comprehensively addressed the decision space. Three examples illustrate how such visualizations differ from the work we have done in visualizing decision spaces that provide option awareness.

Dong and Hayes (2012), for example, developed an “uncertainty visualization” for choosing among alternatives. A primary purpose of Dong and Hayes’ visualization is to enable decision makers to identify when they could benefit from having more information about some key features to reduce uncertainty sufficiently to determine a clearly winning alternative. Dong and Hayes focus on the user-estimated values of the features characterizing each alternative (e.g., the aesthetics of a design, or the reliability of a car), which constitute a special case of the situation space that might be called a “feature space.” Similarly, a fantasy football team prediction system (Miller et al. 2008) displays a feature space consisting of predictions of the statistics that characterize players’ performance. These approaches are different from our work, which enables decision makers to evaluate the consequences of choosing an alternative under a variety of conditions both within and beyond decision makers’ control—thus providing a decision space. Because they do not forecast alternatives’ performance under a broad range of conditions, neither Dong and Hayes’ nor Miller et al.’s visualizations can support a deep exploration of how the features contribute to better and worse outcomes.

In a third example, Hoffman et al. (2006) developed a “roulette wheel” or dartboard-type visualization to show information on the probable outcomes of a particular medical treatment. However, the dartboards can only show a single (or average) outcome for a single set of assumptions. Our research (Drury et al. 2009b) has shown that an effective decision space visualizes the range of outcomes of each option under multiple sets of assumptions, and so is beyond capabilities of Hoffman’s visualization. As we will explain further below, knowing the range of outcomes for each option is essential to gaining option awareness and making improved decisions.

This paper provides for the first time a detailed description of the connection that we are asserting between option awareness and NDM. Along the way, it describes option awareness and team-based (collaborative) option awareness, and presents empirical information to support our assertion that having option awareness can improve NDM.

BACKGROUND

In familiar circumstances, but under time pressure and uncertainty, experienced decision makers employ the simplest form of NDM without comparing any options: they size up the situation and then respond with first option that matches the circumstance (Lipshitz et al., 2001). However, when the relative quality of the different possible courses of action is not obvious, decision makers under NDM begin to use mental simulation to test one option after another to explore the possible results of decisions (Phillips et al., 2004). There are, of course, limits to the variations that can be considered intuitively under emergency time pressures, and even when there are no such pressures (Klein & Brezovic, 1986). Moreover, as the number of viable options becomes overwhelming, unaided decision makers may simply default to the easiest choice to implement rather than make an otherwise satisfactory choice. For example, a study of more than 800,000 people choosing investment fund options for employee 401(k) plans showed that participation rates fell as the number of fund options increased (Sethi-Iyengar et al. 2004). Seventy-five percent of employees participated in their plan when they had two options, but only 61 percent participated in their plan when it had 59 options. The researchers attributed the decreased participation to the employees’ feelings of being overwhelmed. The difficulty in choosing an option from among many alternatives springs from limitations in the brain’s short-term memory capacity (Cantor 2009). A computer-generated display of the decision space offloads this cognitive processing to the computer, which then displays the resulting range of outcomes for each viable option under various plausible environmental conditions. This visualization provides the decision maker with a sort of night-vision goggles for the mind: allowing the decision maker to actually see otherwise obscured relationships between options rather than requiring them to mentally simulate each one. By returning choice to a perceptual comprehension process, we enable decision makers to apply their more powerful visual, pattern matching, recognition capabilities of NDM rather than their more limited capacities for mental simulation.

Before exploring the option awareness—NDM connection further, it is helpful to understand more about these decision space displays that enable this perceptual processing.

Decision Space Visualization

Computer-based forecasting models can assess dozens of options with hundreds or thousands of variations that result from dealing with uncertainty. Uncertainty arises when there are variables outside of the decision makers’ control, which are called exogenous variables. For example, consider the case in which a fire breaks out in an historic building. The chances for successfully dousing the fire with two fire trucks will be much different if high winds arise to fan the fire's flames, versus if a drenching downpour occurs. When uncertain exogenous variables are present, simulation models can run many "what if" variations to determine many plausible outcomes that can occur due to the interaction of a given option and the various plausible values of the variables. The goal is to significantly reduce the mental simulation cognitive load of conceiving and evaluating this boundless array of contingencies (Nadav-Greenberg and Joslyn 2009). Through exploratory modeling (Bankes, 1993; Chandrasekaran, 2005; Chandrasekaran and Goldman, 2007), we provide the decision maker with an ability to compare options in parallel under a whole range of plausible circumstances and ultimately understand the underlying factors that contribute to the outcomes.

Even for a single option, the costs vary depending on situational conditions beyond decision makers' control, such as whether fire trucks can respond quickly or traffic congestion delays their arrival. Thus there is a multidimensional distribution of possible consequences for each option. Each distribution is a function of the uncertainty of the situation space (e.g., how big is the fire) and the uncertainty inherent in the decision option (e.g., what percent of fire trucks will get to the scene and when). Although an optimal plan would generate the highest expected return on investment, under deep uncertainty (Lempert et. al., 2003), where situation and execution uncertainty are irreducible, optimal strategies lose their prescriptive value if they are sensitive to these uncertainties. In other words, selecting an optimal strategy is problematic when there are multiple plausible futures for each option, as is the case in this example. Instead, Chandresekaran (2005) and Chandresekaran and Goldman (2007) suggest shifting from seeking optimality to seeking robustness for planning under deep uncertainty. Robust options result in acceptable outcomes across the broadest swath of plausible futures.

To enable comparisons of disparate options, each with a distribution of disparate outcomes, we mapped each of those outcomes onto a single multiattributed cost metric. First, consequences such as property damage, injury, and death are assigned monetary values (insurance actuarial tables can be used to assign a monetary value to death). Then, the cost of each outcome in our example scenarios (which we call emergency events, or simply events) is computed by summing the resource cost of acting on the option (trucks, people, etc.), the costs of the immediate consequences resulting from that option, and any costs of future consequences that now may occur due to having enacted the option.

We use a frequency format approach to display the results, which means that uncertainty information is displayed in terms of natural frequencies rather than probabilities. Natural frequencies are absolute (non-normalized) frequencies as they result from observing cases that have been representatively sampled from a population (Gigerenzer & Hoffrage, 1995). Hoffrage and Gigerenzer (1998) provided information in the form of either natural frequencies or probabilities to physicians and found that the physicians correctly estimated the positive predictive values of diagnostic tests more than four times as often when they were given the frequency format presentation.

Ibrekk and Morgan (1987) evaluated several methods for communicating uncertain quantities using continuous distributions rather than discrete probabilities. Participants were asked to make a number of estimates regarding upcoming snowfall using a range of nine visualizations that included common displays (e.g., bar chart, pie chart, or a simple number line) as well as six varieties of probability density functions and a cumulative distribution function. Cues in each display type make only certain parameters explicit, so those that had an obvious midpoint were best at communicating the mean snowfall, but often led to significant overestimations of the probability of snowfall being greater than a certain amount. Visual cues available in the pie chart and cumulative distribution function, on the other hand, showed the opposite effect. The recommendation from this study was that a combination of cues providing information referring to both a probability distribution function and a cumulative distribution function would have the greatest chances of successful communication of risk.

However, Ibrekk and Morgan's study was about drawing information from a single distribution, but not about comparing multiple distributions to each other. Our approach is to support robust decision making through the parallel presentation of multiple distributions, so such findings must be extended to identify cues which not only engender an understanding of each distribution, but also facilitate an effective process to compare their relative desirability. We chose box-plots (Tukey, 1977) as a starting point to provide a simple means of comparing the cost distributions of the options. In the study described above, the box-plot provided reliable estimates of the mean, but less than ideal communication of probability density (albeit not the worst among those tested). However, besides their simplicity, box-plots are a common visualization of distributions that typical research participants can be readily trained to read. We further simplified the box-plot visualization by eliminating outlier data points. Future research will be needed to determine a more compelling visualization approach. The basic layout of the box-plot allows for incremental modifications such as shading or additional markers providing additional cues describing the distribution.

In our experiments, the result of the cost evaluations for the range of plausible futures for a given decision option is summarized graphically by a box-plot for that option. Figure 1 shows the box-plots indicating the range of costs for each option in the example scenario. The top and bottom "whiskers" of the box-plot depict the maximum and minimum cost outcomes, respectively. The top and bottom sides of the box show respectively the cost of the 75th and 25th percentiles of outcomes. The line dividing the box indicates the median cost of outcomes.

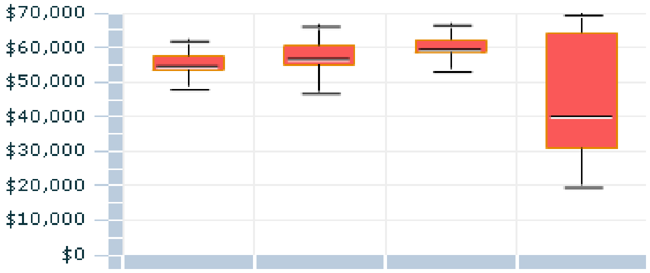


Figure 1. A decision space showing the relative costs of sending between 0 and 3 fire trucks to a fire, assuming no other response.

In our early experiments (Drury et al., 2009a; Pfaff et al., 2010a), participants were given a "best three out of five" rule as a reasonable and easily taught heuristic for comparing distributions and determining the top-ranked option. Applying this rule to the box-plots in Figure 1, sending three fire trucks is considered the top-ranked choice because its corresponding box-plot has the lowest cost for the minimum outcome, the 25th percentile outcome (the lower bound of the box), and the median. Although its 75th percentile (the upper bound of the box) and maximum outcomes are the highest of any option, this option is still best for three out of five of the box-plot parameters and thus is the winner.

Note that the visualization of the options in Figure 1 is presented from the viewpoint of a single fire station, and assumes that no other responders will send assets to handle the emergency. But what if multiple fire stations cooperate, or police and fire plan a joint response? Clearly, collaboration is needed and hence we extended option awareness to assist multiple decision makers.

Collaborative Option Awareness

Extending the example of the fire in the historic building, assume that the fire is reported just as the roads are clogged with bystanders viewing an accident scene. So, three fire trucks are needed because this congestion may delay their arrival, which in turn may allow the fire to grow to a point that requires the three trucks to put it out. Uncertainty about the congestion, uncertainty about the fire's growth, and uncertainty about some future need for firefighting that will possibly be unmet if all three trucks attend this fire, results in the wide range of outcomes for this option. Alternatively, sending fewer trucks will preclude any possibility of the lower cost outcomes that will occur most of the time with this 3-truck option.

Figure 2 shows the decision space for the police, based on the assumption that the police are only concerned with the traffic incident and its local effects. The figure indicates that the most robust option from just the police perspective is to send just one squad car.

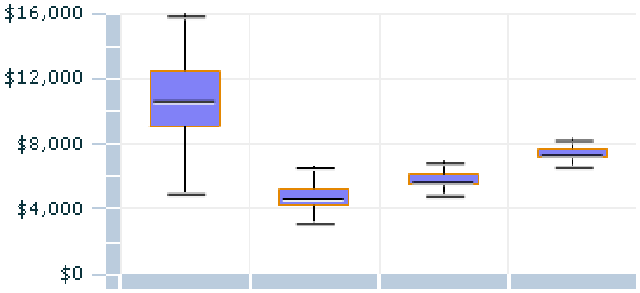


Figure 2. A decision space showing the relative costs of sending between 0 and 3 police cars to a traffic accident.

That perspective, however, ignores the synergy of the two emergency response departments helping each other. For example, it is possible that if the police department sends additional vehicle(s) to clear traffic in favor of the fire trucks, the extra police presence can help the fire trucks reach the fire more quickly. The fire will be smaller upon the trucks' arrival, so fewer trucks would be needed to extinguish the smaller blaze that will cause less damage. Despite the need for more police cars, the total cost to the city could be lower if the cost of sending the extra squad car is less than the reserve value of saved fire trucks (the value of keeping some for future events) plus the resource cost of fire trucks (the cost of sending them to the immediate event).

The effect of this synergy emerges from Figure 3, which illustrates the combined decision space for a collaborative response, and takes the collaborative synergy and the cost tradeoff into account. The most

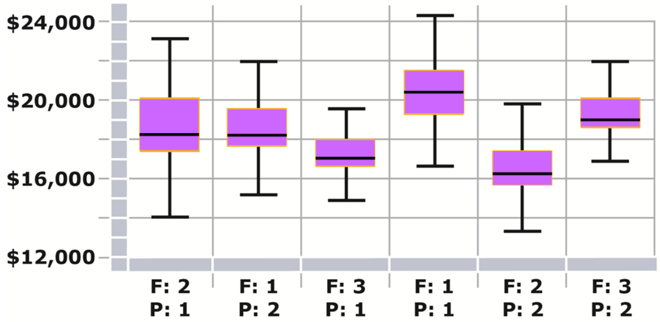


Figure 3. A combined decision space showing the relative costs due to the synergy of sending combinations of fire and police vehicles. Legend: F:x, P:y = x fire trucks and y police cars.

robust combined option revealed by this decision space is to send two fire trucks and two police cars (designated F:2/P:2 in Figure 3). This option wins on four of the box-plot parameters. Moreover, note that considered from the individual decision space perspectives (Figures 1 and 2), combining the costs of the best options appears to yield a total median cost of $45K from those separate views. The combined decision space shows that collaborative synergy results in major costs savings: the most robust option of F:2/P:2 has a combined median cost of only $16.5K, while even the F:3/P:1 option suggested by the individual decision spaces has a median cost of only $17K when synergy is considered in the combined decision space. This is apparently because controlling the traffic, even with one squad car, results in enough time savings to reduce the size and damage of the fire. F:3/P:1 still costs relatively more because more of the costlier fire resources are needed, and that option never will achieve the $13K minimum cost of F:2/P:2.

Our most recent research (discussed in more detail below) suggests that decision makers in the fire and police department would likely be unaware of the level of potential cost savings from this type of cooperation without being able to view a combined decision space. Although emergency responders frequently make tradeoffs in their heads, there are limits to human cognition when analyses involve many variables and high uncertainty, especially when decisions must be made quickly and under stressful conditions. Moreover, if F:3/P:1 was executed, outcome feedback would likely reinforce using the option: it does have a satisfactory outcome distribution. So, unaided by option awareness, F:2/P:2 may never be recognized even as a match for this situation, let alone the most robust option. Our research shows (Liu et al. 2011) that a combined decision space can provide rapid visual comprehension of the likely costs and consequences of collaborative options, leading to better choices.

NDM and Option Awareness

NDM postulates that decision makers look for a match between what they are seeing in the situation space and what they have experienced previously. In other words, decision makers seek to recognize something about the situation that fits within a previously-observed pattern. This tendency to look for what is familiar has been called the recognition heuristic by Gigerenzer and Goldstein (2011).

Our research described in the next section shows that there are patterns to be recognized in the decision space as well. By visualizing the decision space for the user, they can extend to that space the pattern-matching that they otherwise only apply to the situation space. We describe three different experiments that provide empirical support for our assertion of different behavior based on patterns in the decision space.

Ambiguous versus Unambiguous Visual Patterns

One of our experiments (Drury et al., 2009b) introduced conflict patterns into the decision space. A simple or unambiguous pattern in the decision space has no conflicts among the options and one option is dominant in all respects. In a more complex or ambiguous decision space, the pattern of options has conflicts and no clearly dominant option. For example, a conflict might involve one option having a lower median cost, but another alternative having a lower maximum cost. These patterns can be solely in the decision space: that is, the situational context may not vary substantively in two scenarios, but the visual characteristics of their resulting box-plots may be ambiguous in one case but show a clearly dominant option in another case. We hypothesized that unambiguous decisions will be made faster than ambiguous decisions due to the smaller amount of cognitive deliberation that would be needed.

Drury et al. (2009b) used a within-subject design: 20 participants, some from a not-for-profit corporation and some from a university, were presented with both ambiguous and unambiguous box-plot visualizations. Seven of these participants had emergency response experience, and the participants spanned a variety of ages.

All participants were asked to read a paper copy of a one-page introduction to the experiment, which included Institutional Review Board (IRB) information. They were then given a paper copy of a training manual to read and keep as a reference during the experiment as well as a paper copy of Frequently Asked Questions

(FAQ). Next, they were given ten training events in the computerized test bed so that participants could become familiar with its interface and the types of decisions they were being asked to make. After the training, they completed 40 events on the computer test bed during which participants were asked to play the roles of police or fire/rescue commanders.

Each event contained a short textual situation space description of the emergency, the likelihood of another incident occurring soon and a box-plot diagram of the corresponding decision space for comparing the courses of action. Each event was completely independent; what happened in one event did not affect another event and the number of resources available was reset to the maximum for each new event.

During the training events, participants were given feedback. After they entered their estimates for the three parameters (current magnitude, property damage, and potential casualties), they were provided with actual values used in the computational model. After they chose their resource option, they were given the correct number of resources to send based on the model. The test bed was instrumented to capture the amount of time spent making the resource-allocation decision during the test events: declining decision times indicated that the training improved the participants' performance with the interface and the questions being asked of them.

For each of the 40 test events, after reading the textual description, each participant was asked to estimate three parameters: the current magnitude of the emergency incident (via a semantic differential scale implemented as a slider from a low value of 0 to a high value of 7), the likely property damage that could result (radio buttons indicating low, medium or high), and the potential casualties (also low, medium, or high). After setting each parameter, participants were asked to rate their confidence in these estimations using a semantic differential scale from a low of 0 to a high of 7.

Having completed their assessment of the situation space for an event, participants were shown the textual description again, along with the decision space box-plots for that event, and were asked to make a decision regarding the number of resources to send (0 to 5). Immediately after each decision, participants were asked to rate their confidence in that decision on a semantic differential scale (from a low of 0 to a high of 7). All participants were asked two final questions during each event: How much does this decision impact your ability to deal with future situations? (Radio buttons indicated the possible answers of low, medium or high.) What is the likelihood of future situations occurring? (Possible answers were "less than usual," "same as usual," and "more than usual.")

After completing all of the events, participants answered survey questions, including questions probing their subjective assessment of the decision support provided to them.

Results showed that participants' performance was significantly faster when making decisions with unambiguous decision spaces versus the ambiguous ones.

Visual Patterns and Exploration

In our next experiment (Pfaff et al., 2010b), we provided participants with the means to interactively explore the decision spaces via weighting strategies. A limitation of the "best three out of five" heuristic participants were taught to compare options in the earlier experiment is that it assumes an equal weighting of each of the five box-plot distribution parameters. In real-world situations this strategy is not necessarily the best fit in all cases. We could imagine emergency responders concerned about the worst-case scenario choosing options that minimize the maximum cost, e.g., in situations where loss of life seems likely. We termed this weighting scheme emphasize-maximum. Another example is a normalized weighting scheme (called normal) that places the most emphasis on the median cost, and the least emphasis on the maximum and minimum cost outcomes due to the lower likelihood that they will occur. (In this experiment, normal was the default weighting used when the distributions were initially ranked and presented to experiment participants.)

For this experiment, we recruited 41 participants from a major northeastern university and two locations of a not-for-profit corporation. To assess the impact of these new weighting controls, the events and the testing environment from the prior experiment were repeated here. Several improvements in decision performance relative to the prior experiment were observed. When provided with the weighting options, participants picked the highest-ranked option significantly more often and with significantly greater confidence than before.

We hypothesized that the lesser the visual variability of the options in the decision-space visualization, the more weighting strategies participants would consider. We believe that if the options displayed are too visually ambiguous to easily evaluate, the participants engage in more vigorous exploration of the weighting strategies due to a greater desire for confirmatory feedback. The results were consistent with this hypothesis. When there was no completely dominant option in the decision space that could be visually identified as most robust, participants did engage in significantly more exploration of the options. This exploration provided a way for experiment participants to begin learning about patterns in the decision space.

The results from these experiments support our contention that assessing the decision space is largely a perceptual process. We believe that over time, even these “complex” patterns can become as familiar as a complex chess pattern is to a chess expert. Once this learning occurs, applying NDM processes to the decision space can yield efficient and robust decision making.

NDM and Complexity Blindness

In the fire/accident event discussed above, the two most robust options in the individual fire and police decision spaces taken together (three fire and one police) are not the same as the most robust option in the combined decision space (two fire and two police). Because of this difference, we label the combined decision space as conflicted.

The reason for the conflict is because, as noted earlier, the combined decision space takes into account the collaborative synergy and the cost tradeoffs between the collaborators, whereas the individual perspectives do not. Under these conflicted conditions, decision makers that view an event only through their own individual perspective do not have sufficient information in their individual decision spaces or in the situation space to extrapolate the synergies or the tradeoffs. We call this condition complexity blind, a condition that we hope to mitigate through the night-vision goggles of the combined decision space.

To test the effectiveness of this mitigation, in our most recent research, we specifically manipulated the events, creating some with conflicted decision spaces and some with unconflicted ones (where the most robust option in the combined decision space is simply the sum of the most robust options in the individual spaces). In this way we could assess how providing the individual decision spaces, or the combined decision spaces or both, would affect decision making.

In addition, to manipulating confliction, we also manipulated the complexity of the synergy between the collaborating organizations. The synergy in the event described above, between fire and police, is an example of a complex heterogeneous collaboration. Simple collaborations involved only one type of resource (police or fire), but from two different stations. Therefore, their collaboration was merely additive, simply identifying how many resources from which stations would be applied.

Participants were trained and tested as described above, but in this experiment they participated as teams of two. Each participant was in control of one resource (or station), but the team was required to make a combined decision. They were provided with a chat function to plan and coordinate their choices.

Our results showed that in conflicted combined decision space events, when the nature of the synergy between collaborators was also complex, those participants that were provided only with individual decision spaces were able to determine the best or second-best combined option only 30% of the time. However, these complexity-blinded participants were in fact significantly the most confident in their combined choices. Under these same conditions, those receiving only the combined decision space chose the best or second-best option 59% of the time. Those participants that received both decision spaces chose the best or second-best option only 37% of the time. Participants in these latter two conditions had the same level of confidence in their choices, relatively high confidence, but still significantly lower than the complexity-blind condition.

These results provide some insights into findings like those of Shanteau (1992) that even experts perform poorest when situations are complex or novel. The results are also consistent with the boundary conditions for recognition primed decision making (RPD) described by Klein (1998), wherein RPD strategies are less likely to be used with highly combinatorial problems and in cases where the views of different stakeholders have to be taken into account. These are cases where moderately experienced people cannot generate a workable option as the first one considered, but indeed cannot even recognize their own limitations in that regard due to complexity blindness. It is in these cases that intuitive RPD can be extended by the use of decision spaces, providing decision makers with a new view of a decision-event. This view does not require additional mental simulation on their part to consider multiple options over multiple plausible futures, but rather only involves a visual inspection of the space. The visual inspection requires less time to process than having no decision space at all (Drury et al., 2009a; Drury et al., 2009b)

CONCLUSIONS

In this paper, we described how we have extended the basis for naturalistic decision making beyond traditional definitions of situation awareness by providing a computer-generated decision space that enables option awareness. The computer-generated visualization of the decision space enables fast visual comparisons among multiple options simultaneously, which augments mental simulation in RPD even in complex, uncertain settings. By returning processing to perceptual rather than cognitive mental simulation, providing decision spaces and option awareness empirically yields faster, more confident, more robust decisions.

Computer-based forecasting models can assess dozens of options with hundreds or thousands of variations due to uncertainty. Intuitive processing alone cannot adequately process this near-boundless array of contingencies. Through frequency format visualization of exploratory modeling, we provide the decision maker with an ability to compare options under a whole range of plausible circumstances and ultimately understand the underlying factors that contribute to the outcomes. This visualization process facilitates applying NDM to new levels of complexity, which otherwise unfacilitated might have resulted in making simple, easily understood, but wrong choices.

The issues of visually processing this new space need to be addressed. How do we enable less ambiguous visualizations? How do we facilitate agile exploration of this space and mine the underlying data for still covert relationships variables that lead to better and worse outcomes? What are the limits of training people to visually process and accept the implications of this new space? Can these complex patterns become as familiar as a complex chess pattern is to a chess expert?

We believe that addressing these issues can extend NDM to the use of decision spaces, providing decision makers with a new view of a decision-event: a view that frees them from additional mental simulation on their part to consider multiple options over multiple plausible futures.

Future Work

As alluded to above, additional work could be done on making it easier for decision makers to discover and understand the underlying factors and interactions that lead to better and worse outcomes. Our work to date has involved creating the events and designing the synergies according to several formulas (Klein et al., 2011). In the real world, however, synergies will not always be so predictable or formulaic, and decision makers may need extra help in discovering their natures.

We mentioned earlier that we are working on giving decision makers the virtual equivalent of night-vision goggles to see synergies and patterns in the decision space. A challenge remains in presenting the decision space in ways that most effectively aid decision makers in bringing their inherent recognition-primed decision making mechanisms to bear. People are used to extracting patterns from the situation space based on recognizable cues; but what are the most salient cues in the decision space? Much as people are quick to identify human faces that are “off” because they stray outside of the combination of acceptable cues (O’Toole, 2005), we need to better understand the cues that decision makers detect in the artificial landscape that is the decision space. Based on our empirical results, we believe that decision makers are being helped by viewing analyses of the relative costs of plausible options in a frequency format, but we do not know exactly what cues are making this presentation useful.

The experiments accomplished so far presented each event as an isolated situation with one decision point. Except for providing feedback on participants’ assessment of the initial event characteristics during training, we did not provide feedback that would enable participants to learn from their experiences. We do not know if results would differ if participants saw the events unfold in a coherent sequence and were able to use their prior experience with the decision space to influence their future choices. Obviously, this presents an opportunity for future work.

The NDM model assumes that decision makers have relevant expertise based on having encountered similar situations previously. The participants in our experiment had a mixture of backgrounds, with a minority (varying from one experiment to another, but up to approximately one-third) having relevant emergency response experience. In all cases, experience did not affect performance. This result begs several questions. Were the scenarios (which were developed by researchers having emergency response domain expertise) nevertheless so unrealistic that experience would not assist in performance? Or did the decision space visualizations provide a substitute for pattern-matching cues that would be developed through experience, thus leveling the playing field between novices and experts? Or, despite our attempts to isolate the core of the decision making process in a laboratory experiment, is it infeasible to conduct this type of inquiry in a controlled environment? We plan to run our next experiment with emergency responders under controlled conditions, and then implement a field-deployable version of our decision aid that can be investigated during an exercise that at least approaches realistic conditions.

AUTHOR NOTES

The authors would like to thank Sung-Pil Moon and Yikun (Arthur) Liu for their work developing the test bed environments and assistance in running one of the experiments. We would also like to thank the following colleagues at Pennsylvania State University for their earlier contributions to the project: Loretta Moore, Jon Becker, and David Hall. This work was partially funded by MITRE Corporation Innovation Project numbers 43MSR001-EA and 45MSR026-FA. Correspondence regarding this paper should be directed to Gary L. Klein [gklein@mitre.org].

REFERENCES

Bankes, S. (1993). Exploratory modeling for policy analysis. *Operations Research*, *41*(3), (May-June), 435-449.

Cantor, J. (2009). *Conquer Cyber Overload.* Madison, WI: CyberOutlook Press.

Chandrasekaran, B. (2005). From optimal to robust COAs: challenges in providing integrated decision support for simulation-based COA planning. Columbus, OH: Laboratory for AI Research, Ohio State University.

Chandrasekaran, B., & Goldman, M. (2007). Exploring robustness of plans for simulation-based course of action planning: A framework and an example. *Proceedings of the IEEE Symposium on Computational Intelligence in Multicriteria Decision Making*, Honolulu, HI.

Dong, X., & Hayes, C. C. (2012). Uncertainty visualizations: Helping decision makers to become more aware of uncertainty and its implications. *JOURNAL of Cogntive Engineering and Decision Making*, *6*(1), doi 10.1177/1555343411432338

Drury, J. L., Klein, G. L., Pfaff, M. S., & More, L. D. (2009a). Dynamic decision support for emergency responders. *Proceedings of the 2009 IEEE Conference on Technologies for Homeland Security (HST '09)*, 537-544.

Drury, J. L., Pfaff, M. S., More, L. D., & Klein, G. L. (2009b). A principled method of scenario design for testing emergency response decision making. *Proceedings of the 6th International Information Systems for Crisis Response and Management (ISCRAM 2009)*, Göteborg, Sweden.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the 32nd Annual Meeting of the Human Factors and Ergonomics Society*, 97-101.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors 37*(1), 32-64.

Endsley, M. R. (1997). The role of situation awareness in naturalistic decision making. In C.E. Zsambok & G. A. Klein (Eds.), *Naturalistic Decision Making*. Mahwah, NJ: Lawrence Erlbaum, 269-283.

Gigerenzer, G., & Goldstein, D. G. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making*, *6*(1), January 2011, 100-121.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats, *Psychological Review*, *102*(4), 684-704, American Psychological Association.

Hall, D. L., Hellar, B., & McNeese, M. (2007). Rethinking the data overload problem: Closing the gap between situation assessment and decision making. *Proceedings of the 2007 National Symposium on Sensor and Data Fusion (NSSDF) Military Sensing Symposia (MSS)*, McLean, VA.

Hoffman, J. R., Wilkes, M. S., Day, F. C., Bell, D. S., & Higa, J. K. (2006). The roulette wheel: An aid to informed decision making. *PLoS Medicine*, *3*(6): e137. DOI: 10.1371/ journal.pmed.0030137

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, *73*(5), 538-540.

Ibrekk, H., & Morgan, M. G. (1987). Graphical communication of uncertain quantities to nontechnical people. *Risk Analysis*, *7*(4), 519-529.

Klein, G. A. (1998). *Sources of Power: How People Make Decisions*. Cambridge, MA: MIT Press.

Klein, G. A., & Brezovic, C. P. (1986). Design engineers and the design process: Decision strategies and human factors literature. *Proceedings of the Human Factors and Ergonomics Society 30th Annual Meeting*, *2*, 771-775.

Klein, G. L., Drury, J. L., Pfaff, M. S., & More, L. D. (2010). COAction: Enabling collaborative option awareness. *Proceedings of the 15th International Command and Control Research and Technology Symposium (ICCRTS),* Santa Monica, CA.

Klein, G. L., Drury, J. L., & Pfaff, M. S. (2011). COAction: Collaborative Option Awareness Impact on Collaborative Decision Making. *Proceedings of the 2011 IEEE Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA 2011)*, Miami, FL.

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the next one hundred years: New methods for quantitative, long-term policy analysis.* RAND MR-1626. Santa Monica, Calif.: The RAND Corporation.

Lipshitz, R., Klein, G. A., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, *14*, 331-352.

Liu, Y., Moon, S. P., Pfaff, M. S., Drury, J. L., & Klein, G. L. (2011). Collaborative option awareness for emergency response decision making. *Proceedings of the 8th International Conference on Information systems for Crisis Response and Management (ISCRAM 2011),* Lisbon, Portugal, May, 2011.

Miller, S., Kirlik, A., Kosorukoff, A., & Tsai, J. (2008). Supporting Joint Human-Computer Judgment Under Uncertainty. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* *52*(4), 408-412. doi: 10.1177/154193120805200444

Nadav-Greenberg, L., & Joslyn, S. L. (2009). Uncertainty Forecasts Improve Decision Making Among Nonexperts. *Journal of Cognitive Engineering and Decision Making, 3*(3), 209-227. doi: 10.1518/155534309x474460

O'Toole, A. (2005). Psychological and neural perspectives on human face recognition. In S. Z. Li and A. K. Jain, *Handbook of Face Recognition*, New York: Springer Science+Business Media, Inc, 349-370.

Pfaff, M. S., Drury, J. L., Klein, G. L., & More, L. D. (2010a). Decision support for option awareness in complex emergency scenarios. *Proceedings of the 3rd International Conference on Applied Human Factors and Ergonomics (AHFE)*, Miami, FL.

Pfaff, M. S., Drury, J. L., Klein, G. L., More, L. D., Moon, S. P., & Liu, Y. (2010b). Weighing decisions: Aiding emergency response decision making via option awareness. *Proceedings of the 2010 IEEE International Conference on Technologies for Homeland Security* (HST), 251-257.

Phillips, J. K., Klein, G. A., & Sieck, W. R. (2004). Expertise in judgment and decision making: A case for training intuitive decision skills. In D. Koehler & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making*, Malden, MA: Blackwell Publishing Ltd, 297 - 315.

Sethi-Iyengar, S., Huberman, G., & Jiang, W. (2004). How much choice is too much? Contributions to 401(k) retirement plans. In Mitchell, O. S. & Utkus, S. (Eds.) *Pension Design and Structure: New Lessons from Behavioral Finance.* Oxford: Oxford University Press, 83-95.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, *53*, 252-266.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, Mass: Addison-Wesley.

Automated Decision Support On-board Modern Aircraft: A Cognitive Engineering Approach

Giuseppe Montano
University of York
Department of Computer Science

John McDermid
University of York
Department of Computer Science

Paul Cairns
University of York
Department of Computer Science

A novel decision support system for complex fault management procedures on-board modern aircraft is presented. The system is designed on the basis of Cognitive Engineering principles and is aimed at improving pilots' decision-making activity by supporting human cognitive strategies such as mental simulation. Two experiments involving 13 civil pilots are presented. The results show that the framework proposed improves pilots' decision accuracy, decision performance and situation awareness, whilst reducing mental workload and complacency regarding system advisories. In one of the experiments, pilots are provided with probabilistic values representing the degree of uncertainty embedded in the information generated by the system; whilst both high and low degrees of uncertainty enhance the decision making, medium degrees of uncertainty lead to increased decision complexity and mental workload.

KEYWORDS: Decision Support Systems, Situation Awareness, Avionics, Dynamic Reconfiguration, Complacency, Uncertainty

INTRODUCTION

Pilots of modern aircraft are confronted with large volumes of data of diverse nature coming from the on-board instruments. Woods, Patterson, & Roth (2002) investigate the problem of data overload from a Cognitive Systems Engineering (CSE) point of view, giving three different characterizations of data overload: the clutter problem ("too much stuff"), the workload bottleneck (too much data to analyze in the time available) and the problem of finding the significance figuren excessive mental workload and accidents is well documented in the aviation psychology literature (Sarter & Woods, 1994; Sarter, Woods, & Billings, 1997) and in other safety-critical domains such as railways (Johnson & Shea, 2007).

The past two decades of CSE research led to the development of computer technology capable of limiting the impact of mental workload and, more generally, capable of improving the decision making process in safety-critical environments. CSE technology has been developed to improve decision makers' training, to improve forms of communication and coordination, to provide perceptual aids and to enhance the access to relevant data (Smith & Geddes, 2003).

One promising way of supporting pilots during safety-critical decisions on-board modern aircraft is through use of Decision Support Systems (DSS), computer-based technology that is actively involved in the problem-solving and decision-making process.

DSS technology is currently employed with success in a variety of domains, such as railways (Dadashi, Wilson, Sharples, Golightly, & Clarke, 2011), medicine (Shortliffe, Buchanan, & Feigenbaum, 1979), the retail industry (Häubl & Trifts, 2000), nuclear emergencies (Ehrhardt, Päsler-Sauer, Schüle, Benz, & M, 1993; D. Vamanu, Slavnicu, Slavnicu, & Vamanu, 2004), military tactics (Hutchins, Kelly, & Morrison, 1996) and aviation (Davison Reynolds, Kuffner, & Yenson, 2011).

DSS technology can be categorized in a variety of ways. Typical distinguishing features are: framing of the support information, nature of the interaction established with the user, type of knowledge base used, user profiling, algorithms used (see Montano (2012) for a recent taxonomy). From a CSE standpoint, Hollnagel, & Woods (2005) introduced the concept of Joint Cognitive System (JCS): "the combination of human problem solver and the automation and/or technologies that must act as co-agents to achieve goals and objectives in a complex work domain". In this light, this research investigates the effect of novel DSS technology designed on the basis of JCS principles to support the pilot during complex fault management procedures on-board modern aircraft. The technology proposed aims at paralleling certain human cognitive strategies in an effort to enhance the decision making process, e.g., reducing pilots' mental workload, improving their situation awareness.

Supporting Pilots' Cognitive Strategies

Previous Naturalistic Decision Making (NDM) studies revealed the effectiveness of DSS that support cognitive strategies of decision makers during complex decisions characterized by time pressure, high risks and uncertainty (Miller, Wolf, & Thordsen, 1992). Furthermore, a review of the literature shows that mental simulation plays a critical role in the majority of NDM models, such as Recognition-Prime Decision (Klein, 1989), Image Theory (Beach, 1998), Noble's model (Noble, 1993), Explanation-based model (Pennington & Hastie, 1988).

The novel fault management system proposed in this research is designed to automatically generate a number of recovery strategies to mitigate the consequences of an unexpected event, such as a fault or damage to the aircraft. The system also generates decision support information in order to help the pilot in the rapid selection of the best option. The decision support information, specifically conceived to foster mental simulation, includes three types of information which are associated with each decision alternative: (a) explanations, (b) implications and (c) an assessment of the uncertainty embedded in the sensor data used to generate the fault diagnosis. A brief description of each type of information is provided hereinafter, explaining why they are important to support pilots' cognitive strategies.

The explanation of a decision alternative is provided as textual information that answers the question "Why is Alternative A better than Alternative B?" This type of information provides the decision maker with an insight into the logic followed by the DSS in the calculation of the decision alternatives. Studies of human decision behaviour in aviation found that pilots' troubleshooting activity such as fault management, which are part of the global objective of piloting the aircraft, involve the construction of explanations in real-time (Besnard, 2004). Koehler (1991) proposed that firstly, explanations cause changes in the way the problem is perceived by determining which aspects seem to be the more important; secondly, they affect the interpretation of evidence; and, thirdly, they affect the direction and duration of the search. The DSS proposed in this research supports pilots' inferences by automatically calculating this information using a computer-encoded model of the system. Montano (2012) provides all the technical details about the DSS described here.

The second type of information is implication. Intuitively, the implications of a decision alternative answer the question "What are the consequences of choosing Alternative A?" Typically, a fault management decision leads to switching to a degraded operating mode in which only a subset of the aircraft sub-systems are available as a result of a fault or damage. Calculating all the potential consequences of a decision in complex, integrated systems is a difficult and error prone activity that could lead to catastrophic errors of judgment in safety-critical environments. Parasuraman (2000) collates a considerable number of studies that highlight the difficulties humans have with the simulation of the consequences of a course of actions when interacting with a complex automated system. The DSS proposed in this research uses sophisticated Constraint Programming (Tsang, 1993) based algorithms to simulate the implications of a number of potential fault management decisions and shows them to the pilot. A typical implication message for a decision alternative is a list of sub-systems that will be unavailable when the pilot chooses a specific option.

We speculate, and empirically investigate in the two experiments presented later in this paper, that the provision of explanations and implications of each fault management decision alternative would have several benefits on pilots' decision making activity, including a reduction of their mental workload and an improvement of their situation awareness.

The third type of information generated by the DSS proposed here is the uncertainty associated with the fault diagnosis. As discussed later in more detail, the DSS produces a fault diagnosis in real-time using data coming from the network of sensors distributed throughout the aircraft. In a real system, there are circumstances in which the sensor readings are not fully reliable (e.g., one or more sensors are affected by a fault). The DSS proposed here is designed to calculate the uncertainty embedded in the sensor readings and to show this information to the pilot, along with the other decision support information. This should improve the transparency of the system and allow for a more informed decision of the pilot.

The effect of uncertainty on naturalistic decisions is currently subject of investigation in the CSE community. In the weather forecasting domain, Nadav-Greenberg &

Joslyn (2009) investigate the question of whether people in naturalistic settings make better decisions when they have uncertainty information as compared with when they have only a deterministic forecast. Their experimental research indicated enhanced performance with uncertainty information.

Bisantz, Cao, & Jenkins (2011) distinguish between two types of visualization of uncertainty: intrinsic (an integrated component) and extrinsic (an annotation). The authors maintain that, although researchers and designers have developed a variety of methods to represent uncertainty, the investigation of their impact on dynamic decisions has been more limited. The experiments presented in this paper investigate the impact of uncertainty visualized as an extrinsic, numerical annotation associated to a fault diagnosis message (more details are provided later).

Some of the latest research conflicts with earlier studies demonstrating the inability of people to process uncertainty information efficiently (Gilovich & Griffin, 2002; Holyoak, 2005; Tversky & Kahneman, 1973). However, Bisantz et al. (2009) note that previous studies compared human decision making with normative models of rational choice. In the authors' view, a correct assessment of the effect of uncertainty requires the comparison between human decision-making with and without uncertainty information, rather than between human decision-making and normative models.

Building on the latest results from the CSE community, this paper investigates the effects of different degrees of uncertainty on the decision maker behavior. We speculate that both low and high uncertainties have the potential to enhance the decision making performance in critical situations, especially under time pressure, facilitating well-known fast-and-frugal cognitive strategies such as elimination-by-aspect (J. Russo & Dosher, 1983); for instance, high uncertainty of sensor readings associated with a dubious fault recovery suggestion should enable pilots to spot a wrong inference of the system and select another option. More dubious is the effect of medium uncertainty values on pilots' behavior (e.g., 50% of uncertainty embedded in the sensor readings). The second experiment presented in this paper is specifically designed to investigate this issue.

Prior to the description of the experiments, the next section provides a brief overview of the fault management process used in this research: avionics dynamic reconfiguration. This process is currently a debated area of research in both the academic and industrial arenas because of its benefits (both in terms of technology and cost reduction) but also because it could introduce new risks due to the increase of autonomy and authority of the automated logic of next-generation aircraft.

Avionics Dynamic Reconfiguration

The aviation industry is moving towards a new approach to the development of avionic systems: Integrated Modular Avionics (IMA) (Conmy & McDermid, 2001). IMA, in brief, is a term used to describe an airborne real-time computer network consisting of sensors, actuators and a number of computing modules capable of supporting numerous applications of differing criticality levels. The Boeing 787, the Airbus A380, the Lockheed Martin F-22 and the F-35 aircraft all employ IMA technology.

The modularity and flexibility of the IMA architecture enables advantage to be taken of the possibility to reconfigure the avionics to adapt to changing conditions. By pooling the computing resources and allowing them to be shared by different subsystems, at the occurrence of a fault or if the system were to be damaged whilst airborne, the process of IMA Dynamic Reconfiguration (IMA-DR) allows relocation of affected functions to other healthy computing modules.

If a fault or damage affects the computing resources available, an IMA-DR is automatically triggered. A timely reconfiguration decision has to be made, which usually entails choosing which functions should be deactivated because of the degraded operating conditions.

To get an idea of the complexity of the problem, consider that the IMA of the Airbus A380 contains 80 computing modules, each running up to 21 avionics functions that can be activated and deactivated during a reconfiguration (Itier, 2007). The functions in question are inter-dependent, they have different criticality levels which change with the operating conditions, the consequences of deactivating any of them may be very uncertain and, at the same time, the risk is high given the safety-critical context.

Whilst the IMA-DR process cannot be completely automated for safety reasons (Montano & McDermid, 2008), one way of supporting the pilot in this complex type of decision is by providing him or her with the assistance of a DSS.

The following sections present two experiments that are part of a wider series of experiments performed in the context of a four-year long study that examined the issues with high autonomy and authority solutions to the design of dynamically reconfigurable avionics for next-generation aircraft. We first investigated how pilots make decisions during dynamic reconfiguration operations under different operating conditions, including time pressure, heightened stress, different types of decision support information content and framing, and with different cockpit display configurations. We used the results obtained to develop a DSS for IMA-DR, and evaluated it through a series of experiments to assess its effectiveness.

HYPOTHESIS

The following research hypothesis is investigated in this study:

During the process of IMA-DR, decision support information that parallels cognitive strategies and includes explanations, implications and an assessment of the uncertainty associated with the reconfiguration advice provided by the system would have a positive effect on pilots situation awareness, workload, decision accuracy and performance, thus it would improve the overall decision making effectiveness of the pilot and the safety of the process.

Two experiments that address this hypothesis are described and discussed hereinafter. The first experiment focuses on the effects of the explanations and implications on pilots' decision making during IMA-DR. The second experiment goes further and investigates the effects of uncertainty information specifically.

Methods

As part of this research work, we developed the Safe and Interactive Reconfiguration Architecture (SaIRA), a framework for the management of the IMA-DR process based on the Constraint Programming paradigm that (a) generates applicable configurations at run-time by merging information coming from the aircraft sensors, and (b) autonomously generates effective decision support information. For the sake of clarity, SaIRA includes both technology to handle the IMA-DR process (e.g., generating avionics configurations in real-time) and the DSS technology discussed so far in this paper.

For this study, SaIRA was integrated in a flight simulation framework that was used to perform the experiments, and which incorporated eye-tracking technology to help assess pilot performance (Figure 1).

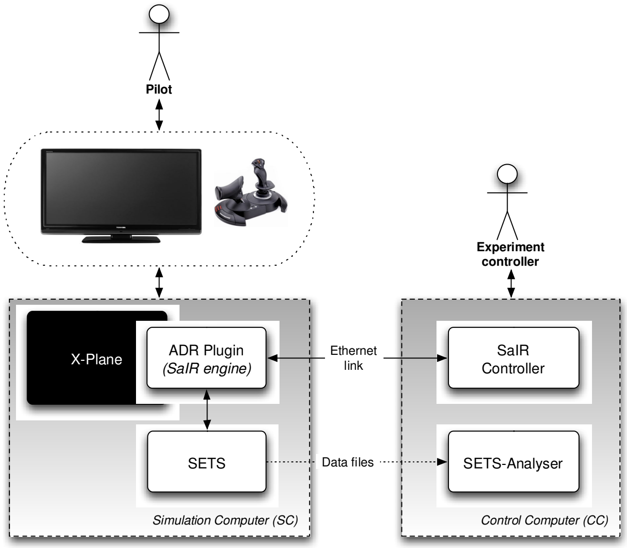


Figure 1. Simulation system architecture used in this study.

Technical details of design and the implementation of SaIRA, including the evaluation of the novel algorithms for automated decision support generation proposed, and about the SaIRA Eye-Tracking System (SETS) are available in Montano (2012).

Thirteen civil pilots from two European airlines, certified to fly the Boeing 737 aircraft, participated in this study. At the time of writing, eleven pilots were resident in the United Kingdom and two in Italy. One of the pilots, of Italian nationality, served as a captain on the B737 and is now in retirement. All pilots were aged between 31 and 68; twelve of them are male, one is female.

Pilots were asked to perform a series of flight simulations in which the operating conditions were purposely manipulated in order to assess the research hypothesis. In a typical scenario, a fault was simulated during a critical maneuver of the flight (e.g., just before landing). A reconfiguration was required to mitigate the effects of the fault and the pilot was required to make a decision about whether to accept the advice of the system or not; in the positive case (s)he also had to choose amongst two or more configuration options to apply amongst those suggested by the system.

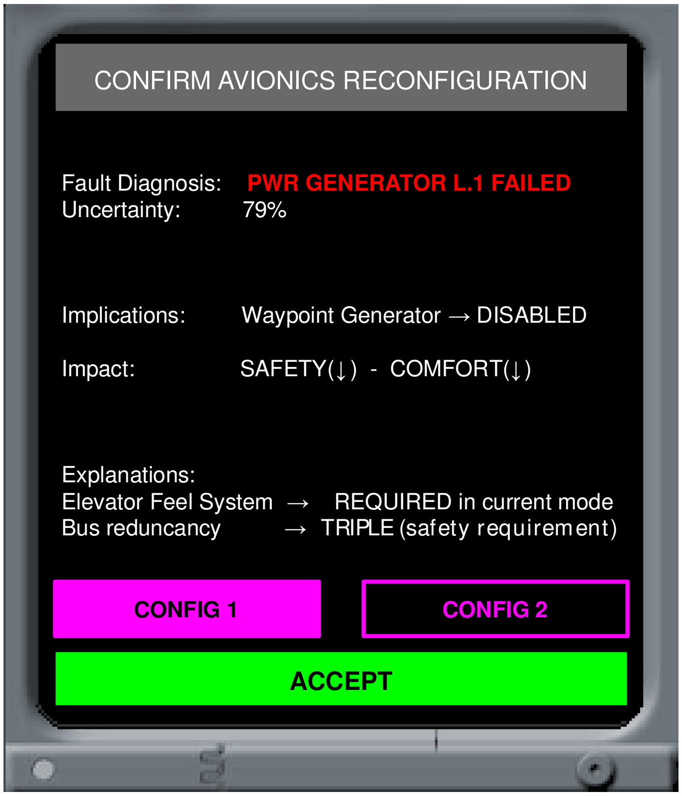
We used two objective and two subjective metrics to characterize pilots' behavior during IMA-DR decisions:

- decision performance (objective): this is a 'composite metric', made up of three sub-metrics: a) decision time, b) decision accuracy, and c) data exploration rate;
- eye-movement (objective): SETS is designed to record a large number of features of the eye

movement. In the two experiments presented here, fixations duration (FD) is taken into consideration, and interpreted as an indication of task difficulty (Rayner, 1998);

- mental workload (subjective): the NASA-TLX (Hart & Staveland, 1988) technique has been used to assess pilots' mental workload (WL);
- situation awareness (subjective): the SA-SWORD (Vidulich & Hughes, 1991) technique was adopted for this study.

In addition, we conducted post-experiment interviews to verify the subjective results.



CONFIRM AVIONICS RECONFIGURATION

Fault Diagnosis: PWR GENERATOR L.1 FAILED
Uncertainty: 79%

Implications: Waypoint Generator → DISABLED

Impact: SAFETY(↓) - COMFORT(↓)

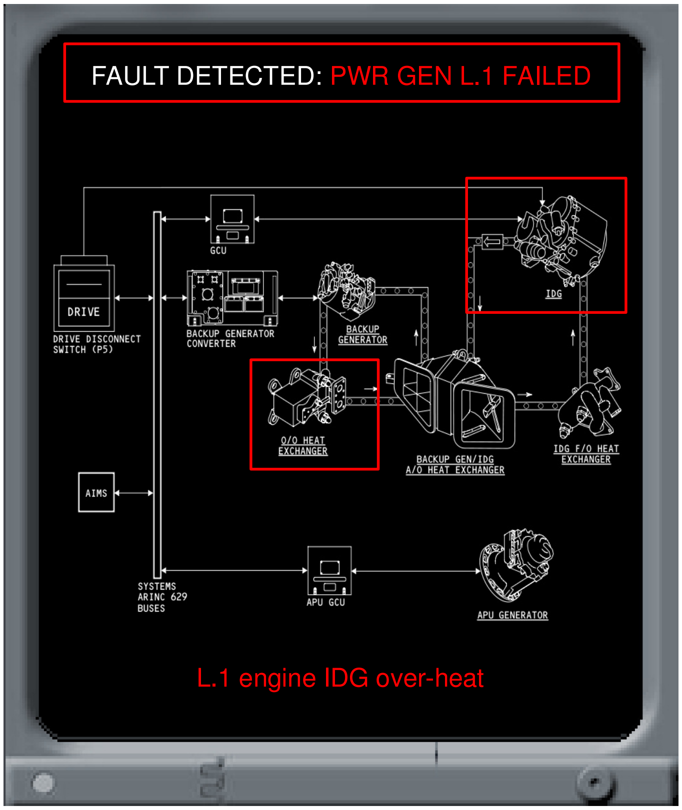
Explanations:
Elevator Feel System → REQUIRED in current mode
Bus reduncancy → TRIPLE (safety requirement)

Figure 2. SaIRA decision support information ('Full SaIRA Information') on the EHSI display.

Figure 2 shows how SaIRA organizes the decision support information on the Electronic Horizontal Situation Indicator (EHSI) of the Boeing 737-900ER cockpit display (used for the simulations). Additionally, schematics about the fault detected by the sensors temporarily replace the content of the Electronic Attitude Director Indicator (EADI) display, as shown in Figure 3.

SaIRA generates the following three cockpit conditions:

- Description only (baseline condition): only 'Fault information' and 'Diagnosis' data is displayed (upper portion of data in Figure 2). The original content of the EADI display is not modified;
- Description & Schematics (controlled condition): EHSI contains the same information as 'Description only' but the EADI shows schematics about the fault detected by SaIRA (Figure 2);
- Full SaIRA Information (controlled condition): full SaIRA decision support information is displayed, including explanations, implications, uncertainty figures and schematics, as shown in Figure 2 and Figure 3.



FAULT DETECTED: PWR GEN L.1 FAILED

L.1 engine IDG over-heat

Figure 3. Schematics that describe the sub-systems mainly affected by the fault.

The eye tracking system superimposes a frame of seven Areas Of Interest (AOI) on the B737 cockpit, as shown in Figure 4; these AOI are used by SETS to characterize pilots' visual attention.

EXPERIMENT A
Description and Aim

Experiment A investigated the effect of explanations, implications and schematics of the fault on pilots' decision-making behavior. The effect of different conditions was examined in terms of decision accuracy, decision performance, frustration, mental workload and situation awareness.

Figure 4. Definition of the AOIs on the cockpit of the Boeing 737-900ER.

In this experiment the effect of the uncertainty embedded in the sensor readings on pilots' decision making activity is not investigated, therefore the fault management information is presented as fully reliable.

Procedure

The pilot was asked to perform six simulations and complete any potential real-time fault management procedure correctly and in the shortest time possible.

Between 30 and 120 seconds after starting the scenario, a fault was simulated and a reconfiguration request was automatically issued.

Two reconfiguration advisories were provided, one of which was evidently wrong (e.g., leading to unsafe conditions). The two advisories were always such that they required choosing between switching off one of two critical functions.

The experiment was structured into three distinct tests. Being a within-subject test, each pilot ran all the simulations:

- INFO_1 (Description only): pilots performed the first two simulations with 'Description only';
- INFO_2 (Description & Schematics): pilots performed the next two simulations with 'Description & Schematics';
- INFO_3 (Full SaIRA Information): pilots performed the last two simulations with 'Full SaIRA Information' (always showing 'FULL reliability', i.e., no uncertainty).

Straight after the last test, both the NASA-TLX and the SA-SWORD questionnaires were given to the pilot.

Expectations

INFO_1 is the baseline condition. As a result of better decision support, we had the following expectations:

- E1: decision accuracy should have progressively improved with INFO_2 and INFO_3;
- E2: it was not possible to make any precise forecast concerning the decision time (DT) when the experiment was designed. On the one hand better decision support should have reduced the time required by pilots to complete the procedure; on the other hand, more information to process could have increased the DT;
- E3: the number of clicks on the reconfiguration buttons should have progressively decreased with INFO_2 and INFO_3. We speculated that the number of times pilots switched from one configuration to another to explore its characteristics would have been indicative of their confusion. Better decision support would have decreased pilots' confusion, hence this value should have decreased, too;
- E4: fixation duration should have progressively decreased with INFO_2 and INFO_3;
- E5: workload should have progressively decreased with INFO_2 and INFO_3;
- E6: frustration should have progressively decreased with INFO_2 and INFO_3;
- E7: situation awareness should have progressively improved with INFO_2 and INFO_3.

Altogether, expectations from E1 to E7 reflect the general expectation of obtaining improved decision performance with INFO_2 and, even more, with INFO_3.

Results

E1: decision accuracy (DA)

Cochran's Q test reveals a statistically significant difference in terms of DA amongst INFO_1, INFO_2 and INFO_3 ($\chi^2(2)=7.091$, $p<0.029$). A pairwise comparison using the continuity-corrected McNemar tests shows that the main improvement over INFO_1 (baseline) is provided by INFO_3. Table 1 contains the descriptive statistics.

Table 1. *Decision accuracy under the effect of different types of decision support information. Columns 'Right' and 'Wrong' contain the number of pilots who made the right or wrong decision respectively.*

	Right	Wrong	Decision Accuracy
INFO_1	15	11	57.69%
INFO_2	20	6	76.92%
INFO_3	22	4	84.61%

E1: decision time (DT)

A significant effect of the type of decision support information on DT is revealed by a Friedman test ($\chi^2(2)=13$, $p<0.02$). A post-hoc test using Wilcoxon Signed Rank tests with Bonferroni correction shows that the stronger decrease of DT is given by INFO_3 ($Z=-2.984$, $p<0.003$). The descriptive statistics are provided in Table 2.

Table 2. *Decision time (in seconds).*

	Decision Time
INFO_1	36.78 (s.d. 6.36)
INFO_2	35.02 (s.d. 5.97)
INFO_3	28.63 (s.d. 8.61)

E3: number of clicks on the reconfiguration buttons (nrCL)

The statistical difference in terms of nrCL amongst the three conditions is confirmed by a Friedman test ($\chi^2(2)=26.297$, $p<0.001$).

As expected, a progressive decrease of nrCL with INFO_2 and INFO_3 with respect to the baseline (INFO_1) is revealed by a post-hoc test performed through a series of Wilcoxon Signed Rank tests (INFO_2 vs. INFO_1: $Z=-3.326$, $p<0.001$; INFO_3 vs. INFO_1: $Z=-3.968$, $p<0.001$; INFO_3 vs. INFO_2: $Z=-2.057$, $p<0.04$). These tests show that the biggest decrease of nrCL with respect to the baseline is provided by INFO_3. The descriptive statistics are provided in Table 3.

Table 3. *Number of clicks on the reconfiguration buttons.*

	nrCL
INFO_1	3.81 (s.d. 1.17)
INFO_2	2.88 (s.d. 1.07)
INFO_3	2.27 (s.d. 0.72)

E3: fixation duration (FD)

A Friedman test reveals a significant influence of the independent variable on the FD ($\chi^2(2)=17.583$, $p<0.01$). The descriptive statistics are shown in Table 4.

Table 4. *Fixation duration (in milliseconds).*

	Fixation duration
INFO_1	410.01 (s.d. 10.55)
INFO_2	379.13 (s.d. 9.32)
INFO_3	354.89 (s.d. 7.04)

The biggest decrease of FD is provided by INFO_3 over INFO_1, as statistically confirmed by the Wilcoxon Signed Rank post-hoc test with Bonferroni correction ($Z=-4.229$, $p<0.001$).

E5 and E6: workload (WL) and frustration (FR)

Table 5 reports the results of the NASA-TLX test.

Table 5. *NASA-TLX results. Parameters: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (PE), Effort (EF), Frustration (FR), Overall Workload (OWL).*

	INFO_1	INFO_2	INFO_3
MD	71.92 (3.42)	63.46 (3.9)	50.00 (4.38)
PD	1.92 (1.21)	1.54 (0.87)	1.15 (0.61)
TD	32.31 (3.47)	27.69 (2.81)	31.15 (3.01)
PE	54.62 (3.94)	58.08 (4.1)	78.85 (2.34)
EF	54.23 (5.71)	46.54 (3.37)	34.23 (3.66)
FR	61.23 (5.72)	52.31 (4.03)	25.00 (2.59)
OWL	52.23 (2.91)	47.15 (1.77)	39.1 (1.83)

A one-way ANOVA test is run on each parameter of the NASA-TLX test. As a result, a strong, significant effect of the independent variable is found on all the parameters except PD and TD (see Table 6).

Table 6. *Results of the one-way ANOVA test on the NASA-TLX results.*

NASA-TLX Parameter	ANOVA result
MD	$F(2,37)=7.95$, $p<0.001$
PD	$F(2,37)=0.171$, n.s.
TD	$F(2,37)=0.597$, n.s.
PE	$F(2,37)=13.605$, $p<0.001$
EF	$F(2,37)=5.32$, $p<0.009$
FR	$F(2,37)=19.219$, $p<0.001$
OWL	$F(2,37)=8.802$, $p<0.001$

The Tukey HSD post-hoc test reveals that INFO_3 provides a stronger improvement then INFO_2 on the baseline INFO_1 (given the number of permutations, for the sake of brevity, the figures are not reported here). Furthermore, a statistical improvement of INFO_3 is confirmed on INFO_2 for PE, FR and OWL.

As one of the parameters of the NASA-TLX method is frustration (FR), this technique also allows collection of data concerning expectation E6. Table 5 and Table 6 reveal that FR decreases statistically with both INFO_2 and INFO_3, confirming the effectiveness of the decision support information produced by SaIRA.

E7: situation awareness (SA)

SA-SWORD does not provide a direct measure of SA but it is designed to give an assessment of which type of information gives the highest SA. As expected, the order for increasing level of SA is (1) INFO_1 (lowest SA), (2) INFO_2, and (3) INFO_3 (highest SA).

A one-way ANOVA test reveals a strong effect of the independent variable on the subjective assessment of SA ($F(2,37)=1860.943$, $p < 0.001$). The Tukey HSD post-hoc test shows that INFO_3 gives the strongest improvement.

DISCUSSION

The main result is that the complete set of decision support information generated by SaIRA (i.e., INFO_3) is very effective in terms of all the dependent variables considered. To a certain extent, DA, nrCL, FD, WL, FR and SA all behaved as expected, providing evidence of a significant improvement in all aspects of pilots' decision experience during DR. An improvement is also found in terms of DT, which we were not in the position to predict.

The improvement brought by a graphical representation of the fault over the baseline, textual information set (i.e., INFO_2 versus INFO_1), is not as strong as in other studies like FAMSS (Hayashi, Huemer, & Lachter, 2006). It must be noted, however, that projects like FAMSS are specifically targeted on the design of effective graphical representations of the fault management information, whilst this study has a different objective: it is mainly tailored to the analysis of the effects of textual information including explanations, implications and reliability information on the interactive fault management process. As the graphical information generated by SaIRA is not as sophisticated and effective as the information produced by more advanced graphic engines like FAMSS, it would be interesting to analyze the combination of the two approaches.

An unexpected result comes from the NASA-TLX: pilots ranked their performance higher in the scale with INFO_3 than with the other information formats. In this regard, Fox & Tversky (1995) argue that feelings of competence occur when people have clear versus ambiguous knowledge. INFO_1 and INFO_2 provide less information than INFO_3, hence there is the possibility that the former two types of information leave room for ambiguities in pilots' minds. With reference to the support theory of reasoning (Tversky & Koehler, 1994), the content of INFO_3 is "unpacked" into more explicit disjunctions, a fact that, according to the theory, increases the "strength of belief" of the decision maker and decreases the ambiguity. We speculate that, as a result of this phenomenon, pilots would feel more competent and thus give themselves a higher performance score.

In the context of a general evaluation of SaIRA, it is particularly important to note the positive effect of INFO_3 on frustration.

EXPERIMENT B
Description and Aim

The textual decision support information generated by SaIRA is made up of explanations, implications and an assessment of the reliability of the reconfiguration advice generated by the system. Experiment A focused on explanations and implications (the information generated by the system was always assumed to be fully reliable); the assessment of the third component requires a different analysis, as shown hereinafter.

On-board modern aircraft, faults are detected and identified by sensor data fusion technology (e.g., 'Block 3.0 avionics' by Lockheed Martin (Caires & Stout, 2002)). SaIRA is designed to calculate the degree of uncertainty embedded in a fault assessment by using algorithms based on Constraint Programming and Evidential Reasoning techniques.

The aim of this experiment was collecting information about the potential effect of different degrees of uncertainty associated with decision support advices of dubious genuineness on pilots' decision behavior.

At this point it is desirable to illustrate some key points about the interpretation of the uncertainty of the sensor information. Sensors are usually characterized in terms of their reliability in the literature; different sensors can have different degrees of reliability with respect to the diagnosis of a specific fault. For instance, a temperature sensor installed on an aircraft engine is more reliable than an on-wing temperature sensor in the assessment of an "engine on fire" event although both sensors can be used for the same diagnosis. However, if their readings are contrasting, the sensor on the engine is considered more reliable (there is less uncertainty in its readings) because it is nearer to the cause of the fault (note that this is a simplistic example).

In the remainder of the paper the terms 'uncertainty' and 'reliability' of the sensor information directed to the pilots are used interchangeably, referring to the same concept but from opposite ends of the same scale: full reliability implies no uncertainty, full uncertainty implies no reliability. This clarification allows for the correct interpretation of the same concept from both the CSE and the sensor network viewpoints.

We advance the following claim:
Providing reliability figures should influence pilots' decision-making performance in the following ways:

- Evidently wrong IMA-DR advisories, when associated with low reliability, would be more easily spotted and avoided than without any reliability figure;
- Low and high reliability options would both be easier to process than medium reliability options, i.e., the decision time would increase with medium reliability options.

Procedure

The pilot was asked to perform three flight simulations and complete any potential real-time fault management procedure correctly and in the shortest time possible. The pilot was also informed that the system could have potentially generated wrong decision support information as a result of technological limitations.

A safety-critical fault was simulated between 30 and 120 seconds after the start. The system was configured to generate only one configuration option; the pilot could either accept it or switch to safe mode.

Pilots were divided in two groups: Group A and Group B. All pilots performed Test 1; then Group A performed Test 2a and Group B performed Test 2b, as follows:

- Test 1 - both Group A and Group B: SaIRA generated right decision support information showing FULL reliability. This was the baseline test, aimed at building up pilots' confidence in the system before providing them with wrong information;
- Test 2a - Group A only: SaIRA generated wrong decision support information showing LOW reliability;
- Test 2b - Group B only: SaIRA generated wrong decision support information showing MEDIUM reliability.

Expectations

The following results were expected:

- E1: workload should have been higher with MEDIUM reliability than with LOW or FULL reliability;
- E2: fixation duration should have been higher with MEDIUM reliability than with LOW or FULL reliability;
- E3: decision time should have been higher with MEDIUM reliability than with LOW or FULL reliability.

Results

This experiment has a mixed factorial design. The two independent variables are the correctness of decision support information (which can be either 'correct' or 'incorrect', with the former being the baseline condition) and its reliability (either 'LOW', 'MEDIUM' or 'FULL', with FULL being the baseline condition). Correctness is the within-subjects independent variable (i.e., all pilots test both its conditions) and reliability is the between-subjects independent variable (i.e., Group A is tested with the 'LOW' reliability condition and Group B is tested with the 'MEDIUM' condition).

For fixation duration and decision time, the main effect of both correctness (C) and reliability (R) is assessed; when ANOVA is used (i.e., for WL), the interaction between correctness and reliability factors is also assessed (CR).

It must be noted that the main objective of this experiment, as previously stated, is investigating the effect of the 'reliability' factor. However, because of the nature of the decision support information, it was not possible to design this experiment without using both correct and incorrect information.

E1: workload (WL)

The results of the NASA-TLX test are show in Table 7. The factor 'Group' tests for the difference of reliability (degREL) whilst the factor 'Test' examines the effect of the correctness of the information provided. Physical demand is not reported because it was rated null by all pilots.

Table 7. *NASA-TLX results.*

	Test 1	Test 2	Group A	Group B
MD	66.67	74.14	64.58	76.25
	(2.07)	(3.83)	(2.71)	(2.83)
TD	35.42	45.42	34.17	46.67
	(4.15)	(4.01)	(3.53)	(4.28)
PE	83.33	59.17	71.25 (3.8)	71.25
	(2.56)	(2.88)		(5.19)
EF	54.25	67.5	57.17	64.58
	(2.85)	(5.06)	(4.39)	(4.46)
FR	25.83	61.67	34.58	52.92
	(2.74)	(7.24)	(2.85)	(9.74)
OWL	43.26	59.25	47.2 (1.51)	43.26
	(2.23)	(3.56)		(2.23)

A two-way split-plot ANOVA test was performed on each parameter of the NASA-TLX test except physical demand (PD). The results for the main effect of correctness (C), reliability (R) and for their interaction (CR) are reported in Table 8, Table 9 and Table 10 respectively.

Table 8. *Main effect of 'correctness' of the decision support information (two-way split-plot ANOVA).*

NASA-TLX Parameter	Effect of 'correctness' (C)
MD	F(1,10)=5.031, p<0.049
TD	F(1,10)=5.294, p<0.044
PE	F(1,10)=40.239, p<0.001
EF	F(1,10)=7.28, p<0.022
FR	F(1,10)=80.742, p<0.001
OWL	F(1,10)=66.87, p<0.001

Table 9. *Main effect of 'reliability' of the decision support information (two-way split-plot ANOVA).*

NASA-TLX Parameter	Effect of 'reliability' (R)
MD	F(1,10)=13.517, p<0.004
TD	F(1,10)=4.556, n.s.
PE	F(1,10)=1.722, n.s.
EF	F(1,10)=1.686, n.s.
FR	F(1,10)=30.062, p<0.001
OWL	F(1,10)=7.026, p<0.024

Table 10. Interaction between 'correctness' and 'reliability' of the decision support information (two-way split-plot ANOVA).

NASA-TLX Parameter	Correctness/reliability interaction (CR)
MD	F(1,10)=6.211, p<0.032
TD	F(1,10)=2.353, n.s.
PE	F(1,10)=1.722, n.s.
EF	F(1,10)=4.941, p<0.05
FR	F(1,10)=44.716, p<0.001
OWL	F(1,10)=51.563, p<0.001

In line with E1, WL with MEDIUM reliability is higher than with the other two cases. It must be noted that WL is higher than the baseline also with LOW reliability.

Interestingly, a peak of temporal demand (TD) is recorded with MEDIUM reliability. This is an unexpected result because no time limits for decisions are set for this experiment. We speculate that the increased perception of TD is a by-product of the increased frustration and cognitive demand. NASA-TLX data was not processed in real-time (as eye movement data), hence it was not possible to question the pilots about this result in their post-experiment interviews.

Another interesting outcome is the negative effect of LOW reliability on pilots' perception of their performance. In practice, the results about their decision accuracy (DA) show that, contrary to the participants' perception, their performance—although lower in average—wasn't statistically worse than in the baseline case (Wilcoxon Signed-Rank test, Z=-1.171, n.s.).

E2: fixation duration (FD)

Table 11 reports the descriptive statistics concerning FD for Experiment E.

Table 11. *Fixation duration (in milliseconds) under the effect of 'correctness of information' (Test 1 vs Test 2) and 'reliability of information' (Group A vs Group B).*

	Fixation duration
Test 1	384.83 (s.d. 10.61)
Test 2	409.53 (s.d. 20.93)
Group A	371.52 (s.d. 8.56)
Group B	421.84 (s.d. 19.86)

The Wilcoxon Signed-Rank test reveals no statistical effect of 'correctness' of decision support information on pilots' FD (Z=0.706, n.s.). Either the pilots did not notice the wrong information (which supports the hypothesis of automation-induced complacent behavior) or they did not have any observable physiological reaction in terms of FD.

On the other hand, the Mann-Whitney U test reveals a strong effect of the 'reliability' factor (Z=2.882, p<0.004); this test compares the Group A and Group B within Test 2. The analysis of FD confirms the increased complexity of processing MEDIUM reliability information.

E3: decision time (DT)

The descriptive statistics for DT are reported in Table 12. Similar results to FD were found for DT. The Wilcoxon Signed-Rank test shows no statistical effect of ‘correctness’ of decision support information on pilots’ DT (Z=1.883, n.s.).

Table 12. *Decision time (in seconds) under the effect of ‘correctness of information’ (Test 1 vs Test 2) and ‘reliability of information’ (Group A vs Group B).*

	Decision Time
Test 1	31.65 (s.d. 2.32)
Test 2	42.85 (s.d. 4.47)
Group A	32.57 (s.d. 1.8)
Group B	41.93 (s.d. 4.89)

The Mann-Whitney U test, instead, reveals a statistically significant effect of the ‘reliability’ factor (Z=2.722, $p<0.006$).

A correlation is found between FD and DT (Spearman’s test: $\rho=0.509$, $p<0.011$), which contributes to the robustness of the conclusions.

DISCUSSION

The main conclusions of this experiment are that (a) MEDIUM reliability worsens DR decision performance and (b) LOW reliability improves pilots’ performance in discarding erroneous information.

In both cases, reliability information has proven to allow pilots to make a more informed decision, which is a determining element in the design of a safety-critical system. This result should not be taken for granted given the evidence from previous studies which revealed how more information is not necessarily better than less (Russo, Schoemaker, & Russo, 1990).

This experiment showed that reliability information has an effect on pilots’ decision performance. An improvement in decision accuracy is detected but it is not possible to draw robust conclusions from this experiment alone because of insufficient statistical power. More robust conclusions about the impact of SaIRA on pilots’ decision accuracy can be obtained from other experiments of our empirical assessment campaign, which will be published in the near future.

CONCLUSIONS

A major contribution of this study is demonstrating the effectiveness of CSE principles in the design of DSS technology capable of improving human decision making performance and accuracy in safety-critical contexts. SaIRA, the novel decision support system proposed in this research, is designed to reflect the joint cognitive system paradigm mentioned in the introduction of this paper.

The main design drivers of many commercial DSS are decision accuracy and decision time: for the designers of this type of DSS, the capability of the system to generate suggestions that lead to the decision alternative with the highest value in the shortest time possible is paramount. Typical examples are automated DSS from the retail domain, such as Amazon. The flight deck provides a peculiar operating environment in which decision accuracy and performance are obviously critical but they are not the only metrics of quality of an effective decision support aid. Because of the special operating conditions of pilots (e.g., time pressure, stress, extreme decision complexity, safety-critical conditions), the capability of the system to be as unobtrusive as possible in terms of mental workload is another equally important quality metrics. Research in the CSE and aviation psychology domains in the last two decades gathered evidence showing that extreme levels of mental workload decrease the human’s ability to react to stimuli and increase the likelihood of human error (Kantowitz & Casper, 1988; Wickens, 2002). By reducing pilots’ mental workload during IMA-DR, SaIRA demonstrates the benefits of applying cognitive engineering principles to the design of next-generation flight deck DSS.

Another important quality metrics for this research is situation awareness. The SA model taken into consideration in this paper is the three-stage model developed by Endsley (1995), which takes into account the representation of an external situation as well as the aims and objectives of the individual involved. Parasuraman, Sheridan, & Wickens (2008) notice that “accurate choice will depend on good SA, but choice is not the same as SA”. In fact, humans make decisions on the basis of the information perceived from the external world, and the way this information is processed changes from person to person. IMA-DR procedures on-board modern aircraft are complicated; in certain circumstances there is no single, best course of action and each pilot can perceive the environment differently and can perform a safe and effective reconfiguration in different ways, depending on his or her objectives, intentions, strategies, values, and other factors. Unlike certain commercial DSS from domains other than aviation, it is extremely important that a decision aid for IMA-DR promotes pilots’ SA so as to favor the correlation of the characteristics of the current situation with similar situations stored in memory and the projection of the possible state of the environment in the near future. SA is particularly important for pilots during real-time, safety-critical decision-making under time pressure and heightened stress. By improving pilots’ SA, once again

SaIRA demonstrates the effectiveness of the CSE approach to produce DSS technology capable of enhancing the safety and effectiveness of modern IMA-DR systems.

Previous research from the cognitive engineering and decision making domain shows that decision makers have difficulties evaluating and processing uncertainty information (see Nadav-Greenberg, Joslyn, & Taing 2008) for a discussion of this topic and the related controversies). In a study on DSS technology for weather forecasting problems, Bisantz et al. (2009) notice that uncertainties are not usually presented to the decision maker as part of the primary information displays because of presumed processing and interpretation difficulties; usually users have to take additional actions, such as selecting specific information sources, to obtain uncertainties. The SaIRA user interface is designed to integrate uncertainty information with fault diagnosis, schematics, explanations and implications, making the full set of decision support information readily available to the pilot all at once. The experiments presented in this paper demonstrate that this approach has benefits on the decision performance with both high and low degrees of uncertainty; medium uncertainties have the by-product of increasing the decision making complexity in situations of time pressure.

In conclusion, SaIRA is found to improve human decision accuracy, decision performance and situation awareness during dynamic reconfiguration decisions on-board modern aircraft. Other ancillary results also cooperate to attest the effectiveness of the decision support framework proposed, e.g., reduced cognitive workload and reduced frustration in situation of heightened stress and time pressure. The positive results obtained in these experiments make the cognitive framework developed in this research a promising approach that we plan to study also in decision-making contexts different to aviation, such as nuclear power plant control.

AUTHOR NOTES

Correspondence regarding this paper should be directed to Giuseppe Montano [g.n.montano@gmail.com].

REFERENCES

Beach, L. (1998). Image theory: Theoretical and empirical foundations. Mahwah, New Jersey London: Lawrence Erlbaum.

Besnard, D. (2004). When mental models go wrong: Co-occurrences in dynamic, critical systems. *International Journal of Human-Computer Studies, 60*(1), 117-128. doi:10.1016/j.ijhcs.2003.09.001.

Bisantz, A. M., Stone, R. T., Pfautz, J., Fouse, A., Farry, M., Roth., E., . . . Thomas, G. (2009). Visual representations of meta-information. *Journal of Cognitive Engineering and Decision Making, 3*(1), 67. doi:10.1518/155534309X433726.

Bisantz, A., Cao, D., & Jenkins, M. (2011). Comparing uncertainty visualizations for a dynamic decision-making task. *Journal of Cognitive Engineering and Decision Making, 5*(3), 277-293. doi:10.1177/1555343411415793.

Caires, G., & Stout, J. (2002). Newest advanced integrated avionics software package flown for first time aboard the F-22 raptor air dominance fighter. Retrieved February 13, 2010, from www.lockheedmartin.com

Conmy, P., & McDermid, J. (2001). High level failure analysis for integrated modular avionics. *Sixth Australian workshop on Safety critical systems and software*, *3,* 13-21. St Lucia, Queensland: ACM Press.

Dadashi, N., Wilson, J. R., Sharples, S., Golightly, D., & Clarke, T. (2011). A framework of data processing for decision making in railway intelligent infrastructure. *2011 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, *1*, 276-283. doi:10.1109/COGSIMA.2011.5753459.

Davison Reynolds, H. J., Kuffner, M. P., & Yenson, S. K. (2011). A field demonstration of the air traffic control tower flight data manager prototype. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55*(1), 61-65. doi:10.1177/1071181311551013.

Ehrhardt, J., Päsler-Sauer, J., Schüle, O., Benz, G., Rafat, M., & Richter, J. (1993). Development of RODOS*, a comprehensive decision support system for nuclear emergencies in europe-An overview. *Radiation Protection, 50*(1), 195.

Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *The Journal of the Human Factors and Ergonomics Society, 37*(1), 32-64.

Fox, C., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics, 110*(3), 585-603.

Gilovich, T., & Griffin, D. (2002). Heuristics and biases: The psychology of intuitive judgement (p. 859). New York, USA: Cambridge University Press.

Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload, 1*, 139-183.

Hayashi, M., Huemer, V., & Lachter, J. (2006). Evaluation of an advanced fault management system display for next generation crewed space vehicles. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, 50*(1), 136-140. San Francisco, USA: Human Factors and Ergonomics Society.

Hollnagel, E., & Woods, D. D. (2005). Joint cognitive systems: Foundations of cognitive systems engineering (p. 223). Boca Raton, FL: CRC Press (Taylor & Francis Group).

Holyoak, K. (2005). The Cambridge handbook of thinking and reasoning. (K. Holyoak & R. Morrison, Eds.) (p. 861). New York, USA: Cambridge University Press.

Hutchins, S. G., Kelly, R. T., & Morrison, J. G. (1996). Decision support for tactical decision making under stress. *Proceedings of the Second International Symposium on Command and Control Research and Technology*. Monterey, CA: Evidence Based Research.

Häubl, G., & Trifts, V. (2000). Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing Science, 19*(1), 4-21. Imforms Online.

Itier, J. B., (2007). A380 integrated modular avionics. *Proceedings of the ARTIST2 meeting on Integrated Modular Avionics*. Rome, Italy: INRIA. Retrieved from http://www.artist-embedded.org/artist

Johnson, C. W., & Shea, C. (2007). A comparison of the role of degraded modes of operation in the causes of accidents in rail and air traffic management. *Proceedings of the 2007 2nd Institution of Engineering and Technology International Conference on System Safety* (pp. 89–94). Citeseer.

Kantowitz, B. & Casper, P. (1988). Human workload in aviation. In E. L. Wiener and D. C.Nagel (Eds.), Human Factors in Aviation (pp. 157-187). San Diego, CA: Academic.

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

Koehler, D. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin, 110*(3), 499-519

Miller, T., Wolf, S., & Thordsen, M. (1992). A decision-centered approach to storyboarding anti-air warfare interfaces (pp. 1-47). San Diego, CA: Contract No. N66001-90-C-6023 for the Naval Command, Control and Ocean Surveillance Center.

Montano, G. (2012). Dynamic reconfiguration of safety-critical systems: Automation and human involvement. PhD Thesis, The University of York, York, United Kingdom.

Montano, G., & McDermid, J. (2008). Human involvement in dynamic reconfiguration of integrated modular avionics. *Proceedings of the IEEE/AIAA 27th Digital Avionics Systems Conference*, 2008. DASC 2008.

Nadav-Greenberg, L., & Joslyn, S. L. (2009). Uncertainty forecasts improve decision making among nonexperts. *Journal of Cognitive Engineering and Decision Making, 3*(3), 209. doi:10.1518/155534309X474460.

Nadav-Greenberg, L., Joslyn, S. L., & Taing, M. U. (2008). The effect of uncertainty visualizations on decision making in weather forecasting. *Journal of Cognitive Engineering and Decision Making, 2*(1), 24. doi:10.1518/155534308X284354.

Noble, D. (1993). A model to support development of situation assessment aids. *Decision making in action: Models and methods, 8*(3), 287-305. Norwood, NJ, USA: Ablex Publishing Corporation.

Parasuraman, R. (2000). Designing automation for human use: Empirical studies and quantitative models. *Ergonomics, 43*(7), 931-951. doi:10.1080/001401300409125.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140. doi:10.1518/155534308X284417.

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 521-533.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372.

Russo, J. E., Schoemaker, P. J. H., & Russo, E. J. (1990). Decision traps: Ten barriers to brilliant decision-making and how to overcome them (1st ed., p. 280). Simon & Schuster.

Russo, J., & Dosher, B. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(4), 676-696.

Sarter, N. B., & Woods, D. D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system. *The International Journal of Aviation Psychology, 4*(1), 1–28. doi:10.1207/s15327108ijap0401.

Sarter, N., Woods, D., & Billings, C. (1997). Automation surprises. *Handbook of Human Factors and Ergonomics, 2*, 1926-1943. Wiley Online Library.

Shortliffe, E. H., Buchanan, B. G., & Feigenbaum, E. A. (1979). Knowledge engineering for medical decision making: A review of computer-based clinical decision aids. *Proceedings of the IEEE, 67*(9). IEEE.

Smith, P. J., & Geddes, N. D. (2003). A cognitive systems engineering approach to the design of decision support systems. *Handbook of Human-Computer Interaction* (pp. 656–675). Lawrence Erlbaum Associates Inc.

Tsang, E. (1993). Foundations of constraint satisfaction. dces.essex.ac.uk. Academic press London.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207–232.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, *101*(4), 547–567.

Vamanu, D., Slavnicu, S., Slavnicu, E., & Vamanu, B. (2004). Decision support systems in nuclear emergencies: A scenario-based comparison of domestic and reference tools. *Radiation protection dosimetry, 112*(2), 209.

Vidulich, M., & Hughes, E. (1991). Testing a subjective metric of situation awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 35*(18), 1307-1311.

Wickens, C. (2002). Situation awareness and workload in aviation. *Current directions in psychological science, 11*(4), 128.

Woods, D., Patterson, E., & Roth, E. (2002). Can we ever escape from data overload? A cognitive systems diagnosis. *Cognition, Technology & Work, 4*(1), 22-36.

The Central Mountain Fire Project: Achieving Cognitive Control During Bushfire Decision Making

Lisa M. Frye
University of Melbourne

Alexander J. Wearing
University of Melbourne
School of Behavioural Science

Previous research has found that fireground commanders who demonstrate superior decision making also report feeling more cognitively in control than their poorer performing peers (McLennan, Pavlou & Omodei, 2005). The aim of the current research was therefore to investigate which factors affect *cognitive load* during bushfire decision-making, and also what skills fireground commanders use to achieve *cognitive control*. Three studies were used for the investigation. Each study used human factors interviews with experienced fireground commanders, either after an Australian bushfire, or after a simulation of that bushfire. We found that variations in the quality and quantity of information, communication, and resources (both too few, and too many) influenced cognitive load, and that fireground commanders' reports of cognitive control were influenced by their ability to deal with competing cognitive demands. Specifically, superior fireground commanders described using heuristics, or metacognitive skills based on experience, to pay attention to competing goals, like: containing the bushfire, versus protecting life and property. We suggest that fireground command is an example of a macrocognitive work system (Klein et al., 2003) that requires goal switching, or trade-offs (Hoffman & Woods, 2011), to achieve cognitive control, particularly in high cognitive load conditions.

KEYWORDS: Bushfire, Decision Making, Macrocognition, Metacognition, Cognitive Load, Cognitive Control.

INTRODUCTION

Human Factors (psychological) issues are important in fireground command and one example of those issues is cognitive control. There have been extensive investigations examining this topic. In particular, McLennan, Pavlou, & Omodei (2005) explored firefighter decision making using helmet-mounted video cameras and visual-cued recall debriefs. They found that superior fire fighters reported feeling more *cognitively in control* than their poorer performing peers, who reported feeling *cognitively overloaded* while making decisions on the fireground. The researchers also found that good cognitive control was not necessarily related to years of experience. In fact, after three years in the job, years of experience did not account for differences in fireground commanders' performance (McLennan, Holgate, & Wearing, 2003). Instead, the researchers suggested that superior fireground commanders use *metacognitive skills* to perform well under pressure (McLennan, Omodei, Holgate, & Wearing, 2004). The current research therefore explores what types of metacognitive skills experienced fireground commanders use while they are making decisions.

Metacognition and Cognitive Control

The term *metacognition* refers to the thoughts that we have about our own thinking processes, and also the steps that we take to manage those thinking processes. For example, writing a shopping list is a metacognitive activity because it shows that we understand the limitations of our own memory, and also that we can manage that limitation (by writing the list) to get the shopping done. Understanding the limitations of our own thinking is referred to as *metacognitive knowledge*, and managing our thinking while we perform a task is called *cognitive control* (Dunlosky & Metcalfe, 2009). Cognitive control is particularly important during complex tasks such as piloting a plane or commanding a bushfire response. In these situations people typically experience high *cognitive loads*, which refers to the amount of information that they need to process in order to perform the task. Valot (2002) used a flight simulator to explore these types of issues with pilots.

In the simulation studies, Valot (2002) found that the pilots' previous flying experiences provided them with metacognitive knowledge, and that they applied that knowledge during the simulator flights to achieve cognitive control. Valot also found that the pilots used metacognitive *heuristics*, or rules of thumb based on experience, to: manage the chronological distribution of tasks, manage risk, manage memory, keep track of highly dynamic activity, and to manage the distribution of cognitive load between themselves and technology.

However, the pilots did not use heuristics (or metacognitive skills) to achieve a perfect flight. In fact, none of the pilots' planned flights matched their actual flights in Valot's (2002) study. Instead, they used metacognitive skills to avoid the dreaded flight (or worst case scenario), and to keep the flights within acceptable tolerance limits. In this way, the pilots managed to achieve their objective (flying safely from A to B), while also balancing competing cognitive demands and avoiding cognitive overload.

Fireground Command – A Macrocognitive Work System

Like pilots, fireground commanders also have to deal with competing cognitive demands. However, whereas some of the pilots in Valot's (2002) study were in a single operator environment (i.e., flying solo), this is rarely the case for fireground commanders. Instead, fireground commanders typically work in a command and control structure with many decision makers, particularly during large-scale bushfire responses. This is an important feature of the decision environment for fireground command, and as such, fireground command is best studied as a macrocognitive work system (Klein, Ross, Moon, Klein, Hoffman, & Hollnagel, 2003).

Klein and colleagues describe *macrocognitive work systems* as those where: decisions are complex and often involve data overload; decisions involve risk, high stakes, and are made under extreme time pressures; goals are ill-defined and multiple goals often conflict; and decisions occur in conditions where few things can be controlled or manipulated (Klein, Ross, Moon, Klein, Hoffman, & Hollnagel, 2003, p. 81). Hoffman and Woods (2011) further suggest that these types of work systems may require decision makers to *trade-off* one goal against another. For example, decision makers may have to tradeoff between efficiency and thoroughness, or between achieving an acute versus a chronic goal.

Furthermore, where macrocognitive work systems involve technology, *cognitive systems engineering* can assist with designing technology that supports rather than hinders decision processes (Militello, Dominquez, Lintern, & Klein; 2009). This is particularly important where a small change in one part of the work system may have a flow on effect (including a negative effect) to other parts of the system. Complex decision environments like fireground command are prone to these types of effects. We must therefore seek first to understand the demands and constraints of the cognitive work that people do in fireground command, preferably by studying domain practitioners (Klein, Ross, Moon, Klein, Hoffman, & Hollnagel, 2003). This was the aim of the Central Mountain Fire Project.

The Central Mountain Fire Project

The Central Mountain Fire project was designed to explore how experienced fireground commanders deal with different types of cognitive demands. We specifically wanted to know:

1. What factors affect *cognitive load* during bushfire decision making?
2. What skills improve *cognitive control* during bushfire decision making?

To explore these issues we developed a case study based on interviews with fireground commanders at the Central Mountain Fire (study 1). The results of that case study were then used to build a command post simulation exercise, and we used the simulation to explore the metacognitive skills of two experienced fireground commanders in a repeated measures experiment (study 2). The experiment was then replicated with another four experienced fireground commanders in a high cognitive load condition (study 3). The three studies in the Central Mountain Fire project are outlined below.

STUDY 1 – THE FIREGROUND (2006)

The aim of the first study was to identify what types of factors increase cognitive load for fireground commanders' when they are making decisions. To achieve this, four fireground commanders were interviewed immediately after their shift at the Central Mountain Fire. The interviews, photographs and maps from that fire were then used to create a detailed case study, which is described below.

Methods

Participants

Four male fireground commanders participated in this study and they all had at least ten years' experience with fighting bushfires (10-50 years). The sample comprised both career (paid) and volunteer (unpaid) firefighters, and the men had all held leadership positions with a rural fire agency. For example, all of the participants had performed the role of Captain or Deputy Captain of their fire brigade at least once. The researchers therefore considered this sample to be representative of the experienced fire agency personnel within the region, and also elsewhere in Australia.

Procedure

The fireground commanders were interviewed using the Human Factors Interview Protocol (HFIP; Omodei, Elliot, Walshe, & Wearing, 2005), which is a tool developed by researchers in the Bushfire Cooperative Research Centre (Bushfire CRC) for interviewing fire agency personnel after their shift at a bushfire. For example, the HFIP was used to interview 120 fire fighters at Australian bushfires between 2003 and 2006 for the Bushfire CRC Safety and Decision Making project. The four interviews used here are from that project.

As well as the interview recordings and transcripts there were also maps and photographs available from the Central Mountain Fire. These data sources captured the complexity of a fireground command environment and, to retain this, a qualitative approach to data analysis was deemed most appropriate. The NVivo8 qualitative analysis software was chosen for the data analysis.

Results

The analysis showed that three main factors influenced reported cognitive load: namely, variations in the quality and quantity of firefighting resources, information, and communication. It is worth noting that cognitive load increased when there were not enough resources, information and communication available, and also when there was too much. For example:

> *'…the span of control he had in the morning was (about 5 trucks), and then they tacked (more) on, and all of a sudden he had 12 trucks screaming and yelling at him what to do…when it all fell to crap, he couldn't keep up.'*

In this case, rather than making the situation easier, additional resources seemed to result in cognitive overload for the fireground commander. Perhaps fireground commanders perform better when they have fewer resources? It seems not. One of the interviewees suggested that his colleague missed an important opportunity to bring the fire under control when he rejected additional resources. In these cases, cognitive overload (and fear of cognitive overload) seemed to influence how resources were deployed at the Central Mountain Fire.

Other examples where resources, information and communication increased fireground commanders' reports of cognitive load include:

- Firefighting appliances breaking down or becoming unavailable during the shift.
- Crews deploying without their fireground commander's knowledge, or to a different location than the commander had intended.
- Fireground commanders relying on a wind change that did not occur until much later than predicted.
- Crews and fireground commanders working from incorrect information, such as wrong grid coordinates given in a situation report.
- Reception problems in mountainous areas that required the use of mobile communication vehicles to relay messages, resulting in communication delays.

These issues are not uncommon at bushfires and they are often resolved without mishap. However, for a period of time they also increase cognitive load for fireground commanders, which is what we were interested in here. In this study, one of the participants also observed that:

> *'…(the fireground commanders) had not had the* ***experience*** *to deal with large amounts of resources … there was a lack of experience to deal with large task force or strike teams, and realizing how you need to keep them moving and keep them active and keep planning ahead. To my way of thinking, both divisions, east and the west, were very much just keeping up as opposed to pre-planning …'*

We were interested in this observation because all of the participants interviewed for the study had more than 10 years' experience, one said that he had over 50 years' experience, and these men were representative of their peers at the Central Mountain Fire. However, like McLennan, Holgate and Wearing (2003), this participant suggests that *years of experience* do not necessarily lead to good decision-making on the fireground.

Conclusions

If *years of experience* are not enough to become an expert, then what skills do fireground commanders need in order to become superior decision-makers? The participant quoted above suggests that particular *types of experience* are important. For example, experience with planning ahead, and with deploying large amounts of resources. In fact, although many firefighters measure their experience in years, they also know that this is not necessarily a good measure of decision making ability. For example, we frequently here firefighters say:

> *"Does he have 10 years' experience or 1 year experience ten times?"*

In saying this, firefighters recognize that not everyone in a fire brigade is deployed to fires (some people perform support functions), and not everyone in a fire brigade is deployed to the same number of fires. Similarly, firefighters from different brigades may deploy to different types of events (e.g., grassfires versus forest fires). In this respect, the *types of experiences* that people have may be more important for fireground command than their years of experience. This is consistent with the findings from Valot's (2002) studies, which suggested that different types of experiences provide experts with metacognitive knowledge that they use during complex tasks to achieve cognitive control. We set out to explore these issues with experienced fireground commanders in a simulation of the Central Mountain Fire.

STUDY 2: COMMAND POST SIMULATION (2007)

The aim of the second study was to conduct a command post simulation exercise of the Central Mountain Fire, and to test what types of metacognitive processes fireground commanders use for achieving cognitive control while they are making decisions. We were interested in how they perform during routine situations, and also in a high cognitive load condition.

Methods

Participants

Two expert male fireground commanders participated in the command post simulation. They were both career (paid) firefighters and had at least ten years' experience within their rural fire agency. The men were naïve to the experimental aims of the study and were told that the purpose of the experiment was to test the suitability of the simulation for use as a training tool.

Procedure

The command post simulation was conducted at the headquarters of a rural fire agency, and was based on the Central Mountain Fire case study described in study 1. Three rooms were used for the experiment, one each for: an Incident Controller (confederate), a West Sector Commander, and an East Sector Commander. As at the original fire, the fireground commanders communicated with each other using field radios.

Networked Fire Chief Program

Researchers used the Networked Fire Chief program (NFC, Omodei, Elliot, Walshe, & Wearing, 2005) to produce a map of the Central Mountain area, and to simulate a fire progressing across it like the original fire. The computer simulation was projected onto a wall in each command post. As with the original fire, the behaviour of the simulated fire responded to characteristics of the environment. For example, the fire would change direction with a wind change, and progress faster through grassland than through heavy vegetation. The behaviour of the simulated fire would also respond to fire fighting activities. For example, researchers could build a bare earth containment line (fire break) in the simulation and the fire would slow at that point. Similarly, researchers could deploy fire tankers to the fire and, provided they had enough water, the tankers could protect houses there. In this way, the computer simulation allowed us to replicate the conditions at the Central Mountain Fire, and to test the decision making of experienced fireground commanders.

Design

This study employed a repeated measures design, so each participant experienced two, one-hour trials of the simulation. For the first trial, each participant was tasked with managing either the West Sector or the East Sector of the bushfire situation, as shown in Figure 1.

Figure 1 shows that the participants performed the sector commander role within a chain of command. That is, they reported to an Incident Controller, who in this case was a confederate fireground commander, and they deployed their fire fighting resources by radioing instructions to a Mobile Communications Vehicle. Mobile Communications Vehicles were used at the original Central Mountain Fire, but in this simulation the researchers performed that role. This arrangement allowed the researchers to implement experimental manipulations.

After the first trial, which was a baseline or normal condition, the participants swapped sectors for the second trial, which involved manipulations to produce a high cognitive load condition. The researchers used manipulations that were based on the issues highlighted in study 1, including:

- The bulldozer broke down part-way through the experiment
- A fire truck was deployed to a wrong location.
- The wind change that was forecast did not occur during the one-hour trial.
- Participants were given a wrong grid coordinate during a situation report.

- The researcher's (MCV) responses to participant's radio communications were randomly delayed.

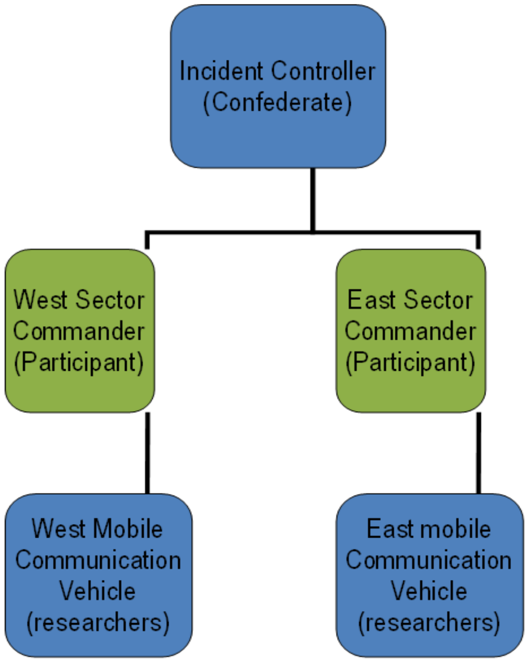


Figure 1. Chain of command for simulation experiment

Visual-Cued Recall Debrief

The Networked Fire Chief (NFC, Omodei, Elliot, Walshe, & Wearing, 2005) program keeps a record of each simulation trial during the experiment. This means that when the participant's trial is finished it can be played back to them so that they can see how the fire situation progressed during their shift. The trial can be replayed in real time or in fast-forward. Importantly for our research, the replay can also be paused so that participants can describe what they were thinking as they made each of their decisions.

For this experiment our visual-cued recall procedure was adapted from the Human Factors Interview Protocol used in study 1 (HFIP; McLennan, Omodei, & Wearing, 2005). Specifically, participants were asked to watch the replay of their simulation trial and to verbalize what they were thinking as they made decisions during the scenario. Like the previous study (study 1), the data for this experiment was mostly qualitative and it was analyzed using the NVivo8 qualitative analysis software.

Results

In all four of the simulation trials the participants described a need to change their plans, and to be adaptive. For example:

> *'…you've got to be able to sit here and think quickly ...work the information that's coming in to you… your strategies are ok, but they're not locked in…you've got to have a fluid environment (and) be able to adjust to the changing conditions.'*

These types of descriptions suggest that, like the pilots in Valot's (2002) study, fireground commanders need to keep track of highly dynamic activity. For the fireground commanders in this study, that meant keeping track of dynamic fire behaviour, and also dynamic human behaviour on the fireground. In fact, one of our participants described this tracking and monitoring process as repetitious:

> *"I must say…the thing that's …yeah repetitious…it's the same things going through your head all the time. It's a continual…an ongoing assessment of…how far (the fire) has progressed, is my plan A, B, and C still valid?"*

This pattern of continual assessment is an important aspect of metacognition. Indeed, Anderson, Oates, Chong, & Perlis (2006) suggest that intelligent systems (like people) use a *metacognitive loop* comprising monitoring and control activities to achieve resilience (or perturbation tolerance) in uncertain conditions. The researcher therefore looked for patterns and repetitions in the visual-cued recall data for this study. One of the most consistent patterns was a reference to two priorities, or two goals. For example:

> *"So, they were the two priorities...get the* ***control line in****, and* ***protect those buildings****…ahead of the fire front."*

Many of the participants' comments related to these two goals, namely: activities that were focused on containing the bushfire, such as getting control lines in; and activities that were focused on protecting life and property, such as protecting individual buildings. The participants also described how they often needed to pursue two goals at the same time:

> *'It might all happen automatically and very quickly, but you've also got the ability in your mind…to do the two things. An ability in your mind to be able to plan worst case and implement best case scenario, at the same time.'*

In this case, the worst-case scenario involved losing houses or people during the fire, including injuries to fire crews. To avoid this, the fireground commanders paid attention to important details in the situation such as the safety of a dozer operator or the location of an individual fire crew. They often described these activities as tactics, or ground truth issues, and sought clarification about them from crews on the fireground.

At the same time, the best-case scenario would be containing the bushfire and preventing it from spreading. To achieve that, the fireground commanders paid attention to the big picture, such as planning where to put containment lines and predicting what the situation would look like in an hour's time. They often described these activities as strategies, or big picture issues.

Switching the focus of attention

The fireground commanders also described a tendency to switch the focus of their attention from one goal (or perspective) to the other throughout the simulation trials. They used several linguistic cues to indicate their pursuit of two goals, and switching between them, for example:

- At the same time…
- In the back of my mind…
- Meanwhile, I'm still concerned about…

Switching the focus of attention is discussed in more detail in the second simulation study (see study 3).

Making decisions automatically (you just do it)

In this study, the fireground commanders also described making a lot of decisions automatically, without really thinking about them. For example:

> *'…decision to deploy a (truck) to property protection, that's easy. Get the dozer out of harm's way and back into a safe area, that's easy. (However) to decide where you're going to cut the trail that needs…you need to look at that…that's a complicated process, where you really need to look at it. But the automatic things (like), property protection, deployment of all your resources, safety of your crews, that sort of thing comes automatically. Yeah, you just do it.'*

Both of the fireground commanders in this study described making decisions this way. However, one of them also pointed out that, while it did reduce cognitive load, making decisions automatically could also lead to human error. For example:

> *'Probably automatically making decisions, it just becomes automated, you just do it. I suppose I don't even think about it. Too automated sometimes (because) you do it naturally and (if) you don't question yourself you can fall into a trap…you can miss something.'*

This participant described being mindful of automatically making decisions, which perhaps accounts for his superior decision making performance during the simulation trials.

Conclusions

In this study the participants described two main strategies for managing cognitive load, namely:

- Switching the focus of their attention between the goal of containing the bushfire and the goal of protecting life and property.
- Making some decisions automatically, without really thinking about it.

These findings suggest that, like the pilots in Valot's (2002) simulation studies, experienced fireground commanders use heuristics (rules of thumb based on experience) to manage cognitive load and competing cognitive demands. The second strategy (making decisions automatically) is also consistent with Klein's (1999) Recognition Primed Decision Model (RPDM), particularly the pattern-matching element. The data also suggests that fireground commanders may experience goal conflict, particularly in the high cognitive load condition. We wanted to explore these issues in more detail with another group of experienced fireground commanders.

STUDY 3: COMMAND POST SIMULATION (2008)

The aim of the third study was to replicate the command post simulation with another sample of experienced fireground commanders. However, this time we were particularly interested in how the participants performed in the high cognitive load condition.

Table 1. *Managing competing cognitive demands in fireground command*

The Big Picture (Contain the Bushfire)	Ground Truth (Protect Life & Property)
(I) wanted more, wanted a bigger picture	Property protection's driving you, or drove me as important.
I need to be focused on what's going to happen over an extended period of time rather than sort of dealing with things in a haphazard manner as they occur in front of me.	at some point in time …the penny dropped that (I needed to) get away from big picture stuff and concentrate now on these individuals.
…a bit hard to divorce myself from the bigger picture or the need for the bigger picture.	…intensely focused on the property protection and those sorts of things.
…without being sort of reacting but being proactive. I've got to take the limited information that I've been presented …and try and fit that into a perception of the bigger picture.	…as they got to the properties (the troops) would get ground truth. They could assess whether they thought the property was safe or hadn't been prepared…our troops are also thinking at that tactical level

Methods

Participants

Four male fireground commanders participated in the third study. As with the previous studies, the participants were all experienced fireground commanders from a rural fire agency. None of them had previously participated in a Networked Fire Chief (NFC) experiment, and like the participants in study two, they were told that the aim of the study was to test the simulation for use as a training tool. They were therefore naïve to the experimental aim of the study, which was to identify how experienced fireground commanders achieve cognitive control during periods of high cognitive load.

Procedure

For this study, we again used the Central Mountain Fire simulation that was described in study 2. However, this time the participants experienced only the high cognitive load condition and expert raters observed their performance.

Results

As with the previous simulation study (study 2), all of the fireground commanders in this study described focusing their attention on two main goals, namely: containing the bushfire, and protecting life and property. In fact, during his visual-cued recall interview one participant neatly summarised how he allocated his attention to these two goals:

> *"…I constantly did this through(out), it was: containment, property (protection), containment, property, containment, property."*

This is consistent with how all of the fireground commanders described their decision making, as shown in table 1 above

As table 1 shows, all of the participants in this study demonstrated *metacognitive knowledge* in that they described their own thinking processes. They also described attempts to achieve cognitive control by switching the focus of their attention, for example:

> *"…and I've got to change my focus and make sure that everyone's been looked after. Now at this stage there's still a tactical focus I suppose, to the point where I'm still thinking about property protection."*

However, changing the focus of attention became more difficult for participants as the fire situation escalated. For example:

> *"…and I suppose with a number of the things that were happening in my mind at the time…it made it difficult for me to divorce myself from that bigger picture and focus more on the immediate matter at hand."*

In a further demonstration of metacognitive ability, participants also described knowing when they were getting it wrong, for example:

> *"There was a realization…that my focus was probably wrong."*

This is consistent with McLennan, Pavlou, and Omodei's (2005) study, in which domain experts were aware of their poor performance. This is important, because it allows experts to take corrective action as the situation progresses.

Cognitive Overload

As the fire situation intensified, all of the participants described feeling cognitively overloaded. This is not surprising because the command post simulation exercise had been tested on previous occasions to ensure that it did induce a high cognitive load. When this happened during the experiment, the participants initially described being unable to switch their attention between the goal of containing the bushfire and the goal of protecting life and property. For example:

> *"..the pressure was on…I had no picture in my head at this point of any containment strategies …and it started to get to me. I don't like not having an idea of containment strategies...I started to actually focus so much on wanting to find some containment strategies that in some ways I lost a bit of focus on the here and now."*

During this study all of the participants described this type of goal conflict. In this case, the participant focussed mainly on developing containment strategies, even though he wasn't having much success. In fact, he seemed to get bogged down with this towards the end of the simulation trial.

Another participant chose the opposite approach. He abandoned the goal of containing the bushfire altogether, and focused instead on the goal of protecting lives and property. He chose this approach because he had done it before during the Canberra bushfires in 2003. This participant received the highest performance ratings from the expert observers in the experiment, and also reported feeling more cognitively in control than the other participants (on a subjective experience questionnaire). However, as the situation escalated further this approach also became problematic, for example:

> *"…started to focus so heavily on what was occurring to the (truck) in here, that I actually lost focus …I got so focused on one (truck) that I lost focus on the other."*

In this case *trading off* (Hoffman & Woods, 2011) the containment goal to pursue only the protection goal seemed to improve the participants' cognitive control. However, he also emphasised that at some point (when conditions were more favourable) he would change his focus again and concentrate all his efforts on getting the fire under control.

Conclusion

Like the fireground commanders in study 2, the participants in this simulation study described focusing on two main goals: containing the bushfire, and protecting lives and property. They also described switching the focus of their attention between these two goals during the simulation trials. However, as the bushfire situation escalated, and cognitive load increased, all of the fireground commanders experienced goal conflict, and found it increasingly difficult to change perspectives.

When this happened, one of the participants abandoned the containment goal altogether, and focused only on protecting lives and property. The remaining three participants described feeling cognitively overloaded while they attempted to manage the situation.

ECOLOGICAL VALIDITY

All of the participants in studies 2 and 3 indicated that the Central Mountain Fire simulation had a high degree of ecological validity, for example:

> *"Very realistic, because you call people (and) they don't answer because they're tied up …or they're out away from the truck or whatever. Or they're on another radio, that's very realistic. So it introduces a level of frustration…that is normal in this sort of situation."*

We also asked participants and expert observers to indicate, on a 9 point scale (with 1 being the lowest score and 9 being the highest score), how the simulation compared with their decision making experiences at real bushfires. In the third study[1], the four participants' responses ranged from 6 to 9 (mean=7), and the four expert observers' responses ranged from 5 to 6.5 (mean=6). It is worth noting that all of the scores were

---

[1] The results were similar for study 2.

above the midpoint in the range of possible scores (5), which the researchers interpreted as an acceptable level of ecological validity for the current experiment. The responses to this item were also consistent with participants' feedback during the visual-cued recall interviews, for example:

> *"Honestly, I think there is a sense of realism to (the simulation). I (can) think (of) a few times in my career where I've been in this situation, where we've (lost) houses and things like that, on a big scale. I've worked in Canberra or recent fires up the coast. It is very similar. It gets to that point where there's a million radio calls coming in and you're just doing whatever you can do while the guys on the ground are just doing whatever they can do to do their best. There's that degree of realism (in the simulation)."*

We also asked the participants and expert observers how useful they thought the Networked Fire Chief simulation was for giving fire commanders an opportunity to examine the strengths and weaknesses of their own decision making skills during uncertainty and time pressure. Again, they provided their responses on a 9 point scale (with 1 being the lowest score and 9 being the highest score). The participant's responses ranged from 7.5 to 9.0 (mean = 8.1) and the responses of expert observers ranged from 6.8 to 7.9 (mean = 7.4). The average of all responses to this item was mean = 7.8. Again, all of the participants' and expert observers' responses were above the midpoint in the range of possible scores (5). It is also worth noting that the participants gave higher scores than the expert observers on both of the ecological validity items. This may suggest that a higher level of psychological fidelity is gained by experiencing the simulation, when compared with observing it.

Possible threat to ecological validity

During the third study there was a possible threat to ecological validity in that one participant described feeling pressured to make decisions, even if they were poor decisions, rather than waiting until he had all of the information that he needed. He attributed this feeling to the presence of expert observers.

The participant's data was retained for this study because fireground commanders frequently have their performances observed, and their decisions critiqued. This was particularly evident during the recent Victorian Bushfires Royal Commission into the 2009 Black Saturday bushfires. Furthermore, pressure to make decisions may reflect a real phenomenon, since another participant also observed:

> *'I suppose to answer your question (about less experienced commanders), if you make a good decision, (then) I don't have to do anything, and if you make a bad decision I (or someone else) can fix a bad decision. (But if) you make no decision at all (then) you kill someone. So I'd rather someone make a bad decision, and then we go 'bad mistake', and then it's corrected. But to stand out there, or in this scenario, and go 'um, um'...that's when you kill someone.'*

According to this participant, making no decision at all represents an *error of omission*, and is more dangerous than making the wrong decision, or an *error of commission*, under pressure. One of the expert observers (present at the time) agreed with him. The authors interpret this claim with caution, since it may reflect a fire culture that favours action (over deliberation), and may apply to particular situations rather than generally. We therefore suggest that more investigation is required.

However, the ability to make decisions quickly, and with incomplete information, may indeed be one of the key competencies required for expertise in fireground command. This would certainly be consistent with Klein's (1999) research about fireground command, and with the subsequent Recognition Primed Decision Model. Furthermore, a *trade-off* favouring efficiency over thoroughness (Bounded cognizance; Hoffman & Woods, 2011) would allow fireground commanders to address one of the key features of macrocognitive work systems, namely extreme time pressures (Klein *et al*).

Another reason that some fireground commanders may favour action (over extensive deliberation) is that action might generate feedback (information) about the situation to aid further decision making (see metacognitive loop; Anderson, Oates, Chong & Perlis, 2006). In this sense, the ability to detect wrong (or imperfect) decisions and correct them may be one of the key metacognitive abilities required for the role of fireground command.

DISCUSSION

The aim of the Central Mountain Fire project was to investigate which factors affect fireground commanders' cognitive load during bushfire decision-making, and also what skills they use to achieve cognitive control.

Cognitive Load

At the Central Mountain Fire (study 1, case study), fireground commanders' reports of cognitive load were influenced by the availability and reliability of resources, information, and communication. The case study was used as the basis for a command post simulation exercise in which resources, information, and communication were manipulated to generate a high cognitive load condition. In all of the studies, cognitive load increased when there were not enough resources, information and communication, and also when there was too much. For example:

- In studies 1 and 3, additional fire fighting resources were rejected because fireground commanders feared an increase in cognitive load.
- In all 3 studies, fire fighters described increased cognitive loads that they attributed to a lack of reliable information, and also to an excess of information.

We also suggest that, for some fireground commanders, the presence of observers increases cognitive load. The studies in the Central Mountain Fire Project explored how experienced fireground commanders achieve cognitive control in these conditions.

Competing Cognitive Demands

During the command post simulation experiments (studies 2 & 3) experienced fireground commanders described focusing their attention on two main goals, namely: containing the bushfire, and protecting lives and property. Protecting life and property seems like an obvious goal for fireground commanders, and it also seems consistent with the goal of containing the fire. After all, the risk to lives and properties drops once the fire is contained. However, a close inspection of the visual-cued recall data shows that when these fireground commanders used the terms *protect life* and *protect property*, they were referring to different types of activities than when they were pursuing *containment* strategies. Furthermore, they described paying attention to different aspects of the fire situation, and they deployed resources in a different way than when they were talking about containment.

Preparing for what will happen next (Containment)

For example, when the participants in these studies talked about containment strategies, they described looking at the big picture, predicting what was going to happen next, linking up road networks to develop a control line, taking big steps, conceding ground, and developing multiple fall back options. As the fire progressed, they also talked about needing more map so that they could develop more control lines. In this way, the containment goal required preparation for a future event (fire impact), and it required fireground commanders to pay attention to longer timeframes. In other words, the pursuit of containment strategies required more *temporal distance* than protecting life and property.

Dealing with the here and now (protecting life & property)

On the other hand, when they talked about *protecting life* and *protecting property*, the participants in these studies talked about taking resources away from containment activities (such as patrolling a control line) and redeploying them to property protection (such as protecting houses). Sometimes this was because patrolling had become too dangerous for fire crews, and at other times it was because the properties or people were under immediate threat. Therefore, in most cases, protection goals were associated with close proximity to the fire front. In other words, the pursuit of protection goals was more immediate, and involved less *temporal distance* than containment activities.

Goal Conflict

During visual-cued recall interviews the participants also described a need to pay attention to both of these goals at the same time. Furthermore, as the bushfire situation escalated, the participants described feeling cognitively overloaded because they were unable to focus on both goals. Consequently, we propose that these two goals represent competing cognitive demands for fireground commanders in this scenario.

Cognitive Control

The fireground commanders in this project achieved cognitive control by switching the focus of their attention from one goal (or perspective) to the other, and also by doing some tasks automatically. However, not surprisingly, participants were unable to maintain this activity under conditions of extremely high cognitive load. In these situations, participants described feeling cognitively overloaded.

During the command post simulation experiments experienced fireground commanders also described making decisions automatically, without really having to think about it. This was particularly evident for tasks like deploying firefighting resources. The descriptions of these types of decisions are consistent with pattern matching, as outlined in Klein's (1999) Recognition Primed Decision Model. This skill was also considered to be lacking among fireground commanders at the Central Mountain Fire (study 1).

Regulating Emotional Responses

In addition to managing cognitive load, participants in the simulation trials (study 2 & 3) also described switching the focus of their attention to regulate their response to losses, including emotional responses. For example, when one participant was asked how he responded to losing a house during the simulation (a failure against the protection goal), he replied:

> *"You have to look at the big picture, we could have lost 10 houses."*

He also described having a similar response at other bushfires in the past. On the other hand, when another participant was asked how he responded to seeing that the entire landscape in his sector had burned (a failure against the containment goal), he replied:

> *"Yeah, but at least we saved those 3 houses in there".*

One participant elaborated on emotional regulation further. He considered it imperative that fireground commanders maintain their composure *'in the face of whatever happens'*, because the fireground commander's response to the situation influences everyone else around them. In these simulations, it seems that changing the focus of attention (from one goal to the other) may provide fireground commanders with a mechanism for achieving that composure (or command presence).

Tradeoffs in Fireground Command

Like the pilots in Valot's (2002) flight simulator, the participants in these simulation studies used rules of thumb based on experience (heuristics) to manage risk. For example, drawing on previous experience, one participant abandoned his goal of containing the bushfire during high cognitive load, and focused instead on protecting lives and properties. In this respect, he also described a trade-off between achieving a long term (chronic) goal and a more immediate (acute) goal. In another example, a fireground commander described abandoning a containment line that had been breached by the fire (acute goal), and falling back to another control line (chronic goal) to improve chances of pulling the fire up. These findings provide initial support for Hoffman and Woods' (2011) notion of bounded responsibility (or acute-chronic goal tradeoff) in macrocognitive work systems[2].

The pilots in Valot's (2002) flight simulator also used metacognitive knowledge to manage the distribution of cognitive load between themselves and technology. Similarly, in our simulation studies some fireground commanders achieved cognitive control by managing the distribution of cognitive load between themselves and other people in the chain of command. For example, one fireground commander described how important it was to recognize his own cognitive limitations, and to use his span of control to keep his tasks in perspective. He described his rule of thumb as: report one level up, manage one level down. In this respect, our simulation studies provide tentative support for Hoffman and Wood's (2011) notion of bounded effectiveness (concentrated-distributed action tradeoff).

Perhaps the most significant support for decision tradeoffs in fireground command is in the notion of bounded perspectives (Hoffman & Woods, 2011). For example, the most consistent pattern in our visual-cued recall data was a tendency for fireground commanders to switch perspectives between: stepping back to see the big picture (usually to pursue containment goals), and zooming in to see the details (usually to pursue protection goals). The current research was exploratory, and it would be worth replicating the studies with another fireground command scenario (such as an urban interface fire), and with different fire response roles (such as Incident Controllers). This would determine whether the findings from the Central Mountain Fire Project can be generalized to other bushfire response scenarios.

CONCLUSIONS

All of the participants in our simulation trials described the protection of life and property as the primary goal of fireground command. At the same time, they all attempted to develop strategies for containing the bushfire, and they switched between these two goals to achieve cognitive control. However, managing these competing cognitive demands became increasingly difficult as the simulation progressed, and participants experienced *goal conflict*, particularly during the high cognitive load condition. In these cases, participants

---

[2] We note that Hoffman and Woods (2011) propose safety as a chronic goal, whereas we found it mostly referred to as an acute goal in this study.

were unable to switch the focus of their attention and they described feeling cognitively overloaded.

Though this research was exploratory, the findings suggest that fire agencies should pay careful attention to the goals prescribed for fireground commanders. This is particularly important in extreme conditions, where decision makers are likely to experience very high cognitive load. These findings also suggest that studying fireground command as a macrocognitive work system (Klein et al., 2003), and using domain experts, has significant value.

We also suggest that further research into the cognitive demands and constraints of fireground command (as well as Incident Management Teams), would inform the development of appropriate decision support systems (and training) for bushfire response. In particular, Hoffman and Wood's (2011) description of the decision space for macrocognitive work systems, in which performance is bound by 5 fundamental *trade-offs*, is worthy of further specific examination.

AUTHOR NOTES

This research was jointly funded by the New South Wales Rural Fire Service (NSW RFS) and the Defence Science and Technology Organisation (DSTO). However, the views expressed here are entirely those of the authors, who gratefully acknowledge the work of Dr. Mary Omodei and Joel Suss (La Trobe University) in developing the command post simulation exercise. Glenn Elliot (La Trobe University) conducted the human factors interviews on the fire ground and, along with Matt Walshe and Peter Hayes (La Trobe University), assisted with simulation trials. Peter Carter (NSW RFS) arranged access to facilities and research participants. Correspondence regarding this paper should be directed to Lisa M. Frye [lisafrye@bigpond.com].

REFERENCES

Anderson, M., Oates, T., Chong, W., & Perlis, D. (2006). The metacognitive loop I: Enhancing reinforcement learning with metacognitive monitoring and control for improved perturbation tolerance. *Journal of Experimental and Theoretical Artificial Intelligence, 18*(3), 387-411.

Dunlosky, J., & Metcalfe, J. (2009). Metacognition (pp. 1-8). Thousand Oaks, CA: Sage Publications.

Hoffman, R., and Woods, D. (2011). Beyond simon's slice: Five fundamental trade-offs that bound the performance of macrocognitive work systems. *IEEE Intelligent Systems, 26*(6), 67-71.

Klein, G. (1999). The recognition primed decision model. In Sources of Power: How People Make Decisions (pp. 15-30). Cambridge, The MIT Press.

Klein, G., Ross, K., Moon, B., Klein, D., Hoffman, R., & Hollnagel, E. (2003). Macrocognition. IEEE Intelligent Systems, 18(3), 81-85.

McLennan, J., Holgate, A., & Wearing, A. (2003). Human information processing aspects of effective emergency incident management decision making. *Proceedings of the Human Factors of Decision Making in Complex Systems Conference*, Dunblane Scotland, September 2003.

McLennan, J., Omodei, M., Holgate, A., & Wearing, A. (2004), Effective fireground incident control decision making: information processing competencies, *Proceedings of the 39th Annual Conference of the Australian Psychological Society*, Sydney, pp. 186-190.

McLennan, J., Pavlou, O., & Omodei, M. (2005). Cognitive control processes discriminate between better versus poorer performance by fire ground commanders. *In H. L. Montgomery, R., Brehmer, B. (Ed.), How professionals make decisions (pp. 209-221).* Mahwah, NJ: Lawrence Erlbaum Associates.

McLennan, J., Reynolds, C., & Omodei, M. (2005). Identifying the causes of unsafe firefighting decisions: A human factors interview protocol. *Bushfire CRC Safety in Decision Making and Behaviour Project* (D2.3). Report Number 1:2005, Melbourne: Bushfire Cooperative Research Centre.

Militello, L., Dominquez, C., Lintern, G., & Klein, G. (2010). The role of cognitive systems engineering in the systems engineering design process. *Systems Engineering, 13*(3), 261-273.

Omodei, M., Elliot, G., Walshe, M., & Wearing, A. (2005). Networked fire chief: A research and training tool that targets the human factors causes of unsafe decision making in wildfires. *Proceedings of the Eighth International Wildland Fire Safety Summit*, April 26-28, 2005 Missoula, MT.

Valot, C. (2002). An ecological approach to metacognitive regulation in the adult. *In P. Chambers, M. Izaute, P. Marescaux (Ed.), Metacognition: Process, Function and Use* (pp. 135-152). Boston, MA: Kluwer Academic Publisher

Improving the Usability of Electronic Health Records through Decision-Centered Design

Christen E. Lopez
Applied Decision Science, LLC

Laura G. Militello
Applied Decision Science, LLC

Nicole B. Arbuckle
University of Dayton Research Institute

Steve P. Wolf
Applied Decision Science, LLC

A his paper describes the application of Cognitive Systems Engineering (CSE) methods to a rapid turnaround usability analysis of five healthcare software modules that leverage electronic health records (EHRs). The Decision-Centered Design (DCD) framework of CSE was modified into a rapid usability process to accommodate project constraints of a tight development cycle and limited access to users and the software. The rapid usability process consisted of three steps: learning about each application, establishing common ground between the users and developers, and articulating recommendations for improvement. Recommendations to the developers were organized in terms of cognitive support requirement, issue class, and suggested implementation timeframe. The rapid usability process was an adequate compromise between the rigors of CSE methods and the constraints placed on the research team.

KEYWORDS: electronic health records, decision-centered design, cognitive task analysis, time-based display, usability

INTRODUCTION

The usability of a software application is an important aspect to consider throughout the development cycle. Ideally, user feedback is incorporated throughout the design and programming process so that the result is a product that the end users will like and use. However, a user-centered design cycle is not always implemented due to various constraints, and usability analysis can sometimes occur at the very end of the cycle, just prior to release. Traditional usability analysis focuses on whether the software helps users accomplish required task performance, minimizes training needs, is reliable, and standardized (Shneiderman & Plaisant, 2005). While these are all important factors, we believe there is an opportunity to reframe this traditional usability approach to incorporate concepts and themes from the naturalistic decision making field. In particular, we assert that by putting the decisions that users will need to make with the software at the center of the analysis, the end product will not only be usable, but will also better support users' decision making and expertise.

In this paper, we describe how we applied a decision-centered design (DCD) framework to the evaluation of five electronic health record (EHR) applications just prior to their release in the Indian Health Service's facilities. DCD is one of five Cognitive Systems Engineering (CSE) frameworks, each of which is tailored towards addressing the challenge of designing technology, training, and processes to manage cognitive complexity (e.g., decision making, judgment, problem solving) in sociotechnical systems (Militello, Dominguez, Lintern, & Klein, 2009). Sociotechnical systems are characterized by extensive collaboration and links between humans and technologies. CSE aims to support users in these complex environments and support users in managing the cognitive complexity.

The five main frameworks of CSE include Cognitive Work Analysis, Situation Awareness-Oriented Design, Work-Centered Design, Applied Cognitive Work Analysis, and Decision-Centered Design. DCD differs from the other four CSE frameworks because of the emphasis it places on exploring and using the key decisions in high risk, time-pressured situations as a means to support human cognition. DCD methods focus on eliciting the critical and difficult decisions (as opposed to all possible decisions) that need to be supported and designing technology to support users in handling those decisions. A key assumption of DCD is that by designing for the most challenging use cases, everyday use cases will be addressed along the way (Militello & Klein, in press).

DCD theory evolved from the practical application of Naturalistic Decision Making (NDM) models and methods (Hutton, Miller, & Thordsen, 2003; Crandall, Klein, & Hoffman, 2006). There are five phases of DCD (Militello & Klein, in press; Crandall, Klein, & Hoffman, 2006):

- Preparation: learn about and understand the task domain and users
- Knowledge elicitation: apply CTA methods to understand key decisions
- Analysis and representation: identify central issues and decision requirements
- Application design: support user decision making by translating decision requirements into design elements
- Evaluation: evaluate the system in terms of supporting the user.

However, there are no established frameworks for conducting a decision-centered usability analysis. For the analysis described in this paper, we used the key tenets of DCD to develop an approach that uses cognitive support requirements that are congruent with what is understood about how people make decisions in the real world. The resulting framework is a straightforward and simple method which can be effectively applied despite common constraints (such as aggressive development and release cycles, limited access to the software and users). Using EHR technology as a test bed for this approach was appropriate because it is an emerging centerpiece of many of the next generation healthcare software tools. Additionally, the healthcare environment is characterized by demanding elements, such as high stakes and time pressure, thus making a DCD approach that focuses on supporting user cognitions in such complexity valuable.

Indian Health Service (IHS), an agency within the Department of Health and Human Services, has had an electronic health information system in place since the early 1980s, providing powerful tracking, documentation, and decision support capabilities helping to deliver comprehensive healthcare to approximately 1.9 million American Indians and Alaska natives nationwide. The Resource and Patient Management System (RPMS) has allowed IHS to become a leader in health information technology. In 2005, IHS incorporated a graphical user interface (GUI) front end, RPMS EHR. They are in the midst of a transition from the old "roll and scroll" (i.e., text-based) interface to a "point and click" graphical interface with several of the ancillary RPMS applications. This means that currently, some of the applications in RPMS are accessed through the text-based interface, and others are accessed through GUIs.

The research team analyzed the usability of five GUI-based applications that were about to be implemented within the RPMS suite of clinical and practice management software (patient registration, scheduling, admission/discharge/ transfer modules, medical event tracking, and clinical flowsheets). There were several objectives to this project:

- complete rapid turnaround usability analysis to fit the development and release schedule;
- apply NDM concepts and theory to structure design improvement recommendations to support decision making and other cognitively complex tasks;
- positively influence future RPMS application designs to continue supporting cognitive needs of the users in IHS facilities

The ultimate goal of this analysis was to apply CSE methods to create design recommendations that would support human expertise and decision-making.

Methods

Methods for this project were largely driven by three key constraints, none of which are uncommon in the world of software usability analysis. The first constraint was that the sponsor requested feedback for each application within ten days of an initial structured walkthrough of the application. As a result, the usability analysis needed to move very quickly. The second constraint was that the development teams, the users, and the usability team were geographically distributed. Given the project timeline, there would be no opportunity for face-to-face meetings, interviews, or observations in clinical settings. The third constraint was that current prototypes of the applications were not available to the usability team for in-depth exploration and first-hand interactions. The development teams provided live demonstrations via web conferencing, and then followed up with screenshots to be used in discussions with users and as a basis for communicating recommendations. These constraints called for innovative adaptations of traditional usability methods, described below.

The research team applied a rapid-turnaround usability process (Figure 1) that consisted of three main steps: (1) getting up to speed through structured walkthroughs of the software and discussions with the development team; (2) establishing common ground between the users and

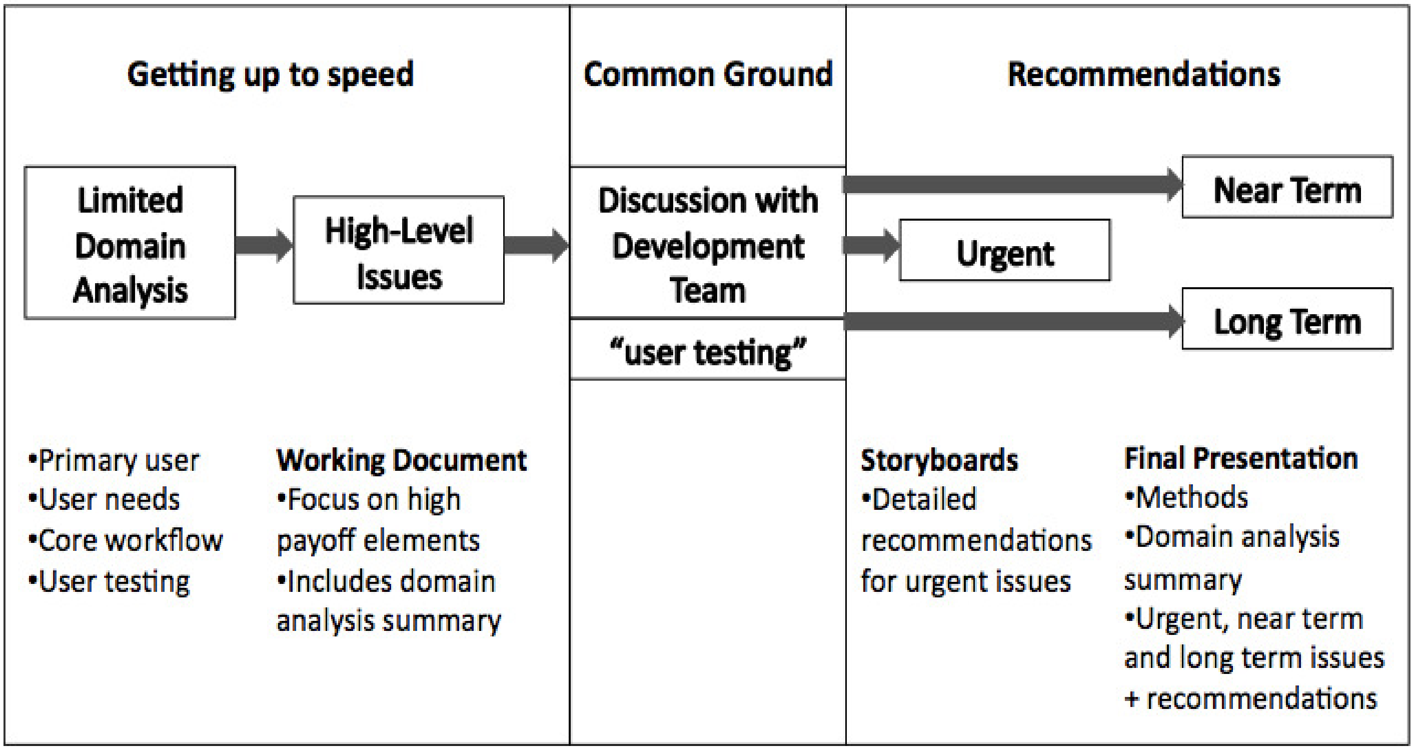


Figure 1. **Rapid usability process**

the developers; and (3) organizing recommendations for software improvements according to cognitive support requirements, issue type, and implementation timeframe.

In line with the first phase of the DCD process, preparation (Militello & Klein, in press), the first step in our rapid-turnaround usability process was to gain an understanding of what the software application was supposed to do and how it worked. After attending a structured walkthrough of the most recent release, the usability team reviewed requirement documents, learned about pre-existing versions (either text-based or previous GUIs), and became familiar with the interface through screenshots and a review of existing application documentation including user and technical manuals.

The second step, establishing common ground, had two main components. The first was to conduct user interviews that leveraged cognitive task analysis elements (similar to the knowledge elicitation phase in traditional DCD). Interviews occurred via web conferencing and conference lines (users were located in IHS and Tribally-operated facilities across the country).

The usability team interviewed 4-6 clinical users who were most likely to be familiar with the existing procedures each application was designed to replace. Interview sessions were broken down into three parts: background information about the participant, a walkthrough of static screenshots and discussion about how the user would accomplish tasks using the application, then a general discussion about overall feedback and any questions about the application. Researchers recorded reactions, comments, questions, expectations, and suggestions. The second component of establishing common ground consisted of informally sharing the interviewees' feedback with the developers. Because the usability team had limited exposure to each application and had not been involved in the design phases, it was important to understand user reactions in the context of previous design decisions, trade-offs considered, and overall goals for each application.

The third step of the rapid turnaround usability process was a formal presentation of recommendations in both a web conference briefing and report format. Recommendations were framed in terms of cognitive support requirements. The intent was to help the development teams think beyond individual elements (i.e., make the font bigger) to the impact of each design element on the user. Through the application of this process, a usability framework compatible with decision-centered design (detailed below) emerged.

Finally, because the development of RPMS GUIs in IHS is an iterative process, it is important to maintain the lessons learned from this analysis. As new technologies become available and the nature of clinical work continues to evolve, the applications reviewed in this project will likely be replaced by newer versions in the

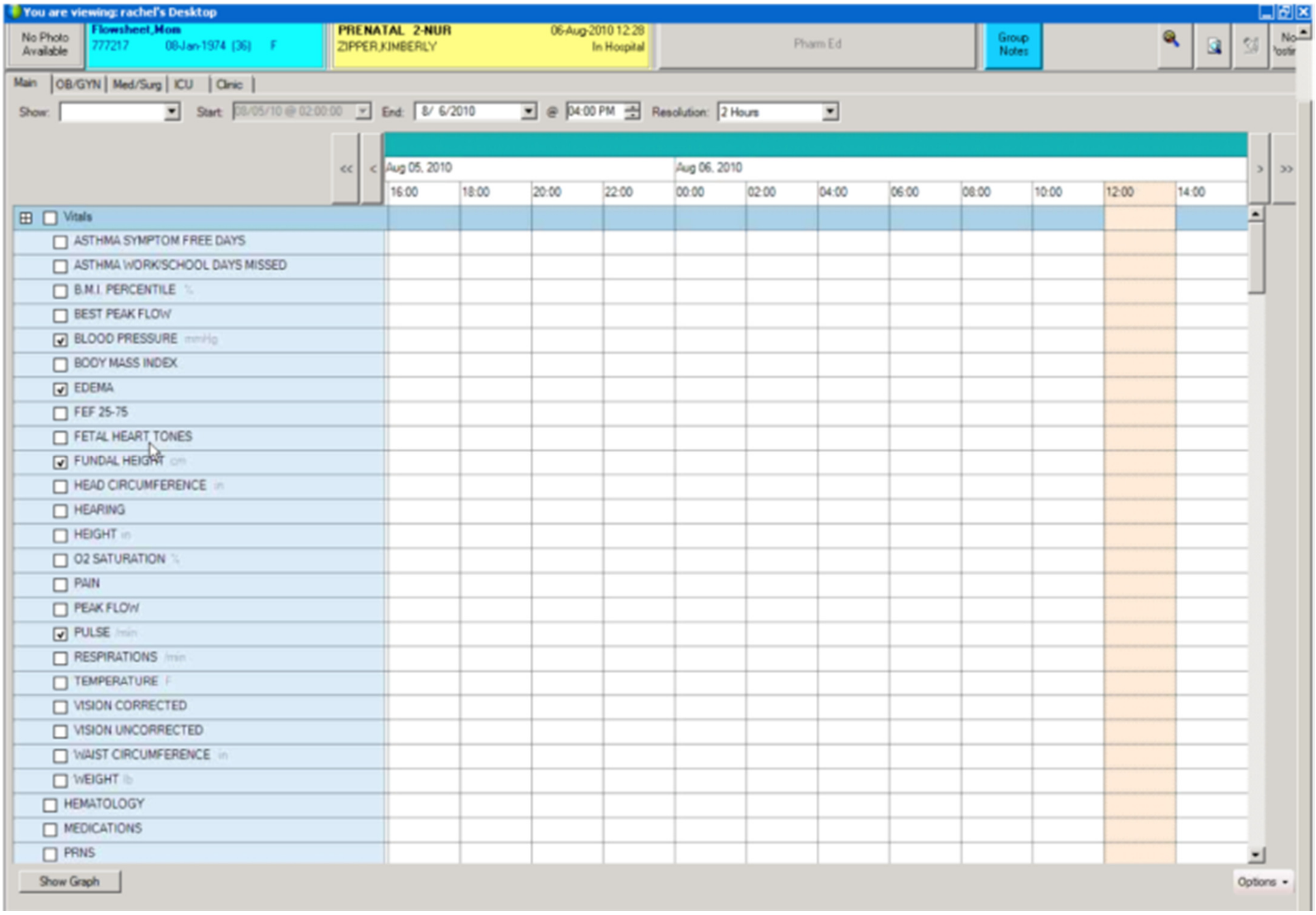


Figure 2. Clinical Flowsheet main view with a fixed grid

foreseeable future. Therefore, the research team compiled a set of usability guidelines that IHS could use in the design of new applications. This document describes the NDM principles and the DCD framework used in this analysis, and will be available much earlier in the design process than previously. This document not only laid out basic usability standards such as font size, color scheme, and window layout, but also ways to improve the software's use as a decision-support tool.

DECISION-CENTERED DESIGN USABILITY FRAMEWORK

After analyzing the data gathered from user interviews, the usability team organized the issues that were raised in terms of cognitive support requirements, the type of risk each issue could introduce if left unaddressed, and a suggested timeframe for implementing recommended changes.

Cognitive Support Requirements

When CSE principles are applied to design work, issues such as information representation, salience of important pieces of information, concepts that are meaningful to the users, relationships, goals, and information flow come to the forefront (Militello et al., 2009). DCD methods in particular are designed to elicit these key issues in terms of the difficult decisions that the users must make in time-pressured environments such as healthcare. As we analyzed the qualitative data from the user interview sessions for the cognitively complex elements that could be better supported (the analysis and representation phase of DCD), a set of six cognitive support requirements emerged. Each requirement addressed a specific aspect of supporting users' cognitive needs.

In line with the overarching goal of the analysis to apply CSE methods to support users' expertise and decision making, recommendations were developed to help users learn the software quickly and accurately, increase ease of use and adoption, reduce common errors (i.e., data entry errors), and increase detection of anomalies. Recommendations were then generated to address the issues unveiled during the user interviews. The cognitive support requirements were used to categorize issues and recommendations in each presentation and report.

Mental Model Support

The users of EHR applications, by definition, work in highly dynamic environments with many complexities. One tenet of the DCD approach is to promote users' cognitive strengths and support their cognitive weaknesses (Hutton, Miller, & Thordsen, 2003). One way the research team incorporated this principle was by framing many recommendations that would help support users' existing mental models, as well as build new ones. The goal was to help users learn the software quickly and thoroughly, since healthcare environments usually do not allow for extensive and repetitive training.

Decision Support

Another way that the research team incorporated this principle into our approach was to emphasize effective decision support as a cognitive support requirement. Healthcare workers must make many decisions through the course of their shifts, many of which have a direct impact on the care that patients receive. The research team therefore made many recommendations to leverage technology's capabilities to help users make quick and accurate decisions (i.e., identify and highlight anomalies)

Error Reduction

Preventing errors is one of the eight "golden rules of interface design" (Shneiderman & Plaisant, 2005). Users need to be able to trust that the application is accurate and reliable. Since errors in healthcare can be especially costly, EHR software should be designed to reduce the likelihood of errors occurring. The usability team made several recommendations to increase error reduction strategies, such as alerting users to anomalies and potential errors.

Perceived Affordances

The term *affordances* refers to the easily discoverable actions associated with a specific object. From the perspective of the user, what does it look like I should do with this? In software design, creating perceived affordances helps guide the user to complete certain actions (Norman, 1999). The effective use of perceived affordances can help users learn and explore the software, thus helping them to quickly develop a strong mental model of the application.

Scanning facilitation

Another general usability guideline is that a screen should not be too cluttered or difficult to navigate, thereby not taxing users' short-term memory (Shneiderman & Plaisant, 2005). To support users when completing tasks in highly dynamic environments, EHR GUIs should facilitate scanning for pertinent information. The information should also be organized to be aligned with specific user tasks and promote information sharing.

Perceived Benefit

One challenge to creating effective software for a group of people to use is to have a balanced cost-benefit ratio from the perspective of the user (Grudin, 1994). In terms of an EHR, users should perceive that the application will make their jobs easier, more effective, and safer than existing methods, while balancing any additional tasks that using the software may entail. To be accepted, it is important that users don't perceive these extra tasks as added duties imposed on them by administrators, with little benefit to their own jobs.

Prioritization

Given the aggressive timeline each development team was working toward, it was important that the usability team provided guidance regarding the importance and potential impact of each issue/recommendation. Each development team needed to determine which recommendations were feasible for implementation immediately and which could be implemented in later releases. Therefore, recommendations were prioritized based on the type of risk each issue might present if left unaddressed. There were three issue classes:

- Type 1- issues that could create individual error risks; these issues could introduce a specific health risk to a patient;
- Type 2- issues that could create aggregate error risks; these issues could introduce error through cumulative effects;
- Type 3- issues that could lead to adoption and long-term use risks; these are issues with the software that will negatively affect user acceptance.

In addition to this classification of issues, recommendations were given an implementation timeframe: urgent, near term, or long term. The following is an example of an issue and the accompanying recommendations.

Many of the applications reviewed were new GUIs based on an existing application in the text-based version of RPMS. Some applications were brand new, and replaced paper-based methods, such as the Clinical Flowsheet application. Nurses traditionally use paper flowsheets to track patient vitals, drug administrations, and note details of which other members of the healthcare team should be aware.

The Clinical Flowsheet application was designed to mimic paper-based flowsheets, with a standard grid format (Figure 2). This grid was more flexible than paper sheets, since users could change the time span to see more or less data. However, if the time span was set too broadly, data would be concealed, which concerned users who were interviewed. Each box would only show the most recent data measurement; so if a box represented 30 minutes, there may be just one measurement. But if the time span was set so each box represented 12 hours, it could potentially hide several data measurements (for example, if rounds were completed every four hours, two additional sets of data would be concealed, while only the third showed). If a box contained more than one measurement, a triangle would appear in the corner that cued users that additional values were available (Figure 3).

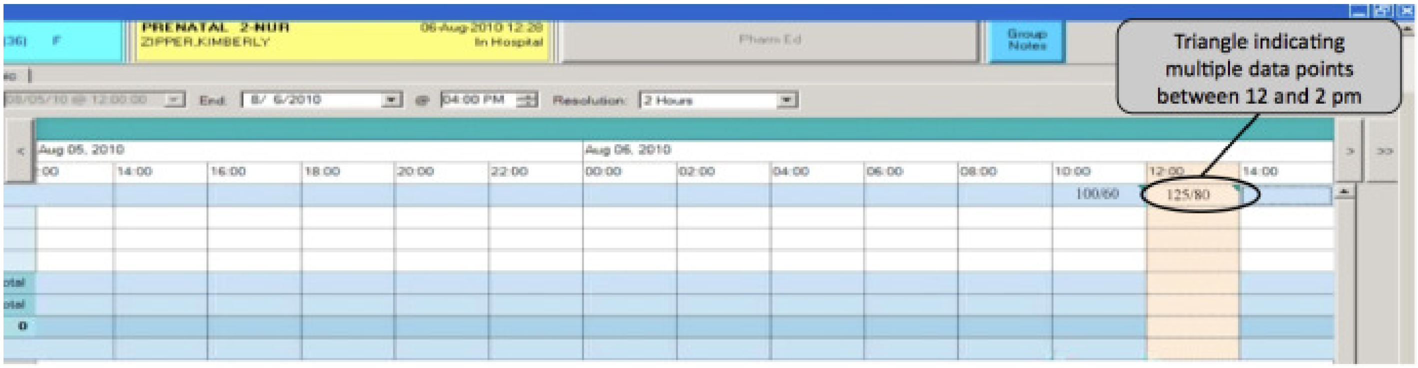
The usability team recommended leveraging technology's benefits over paper to create a display that could better support users' decision making. One suggestion was to create a timeline display that would make trends more visible. For example, if a patient's condition deteriorated unexpectedly, causing a flurry of actions within a short amount of time (i.e., 12 minutes), then stabilized, the original fixed-grid display could hide or misrepresent this data. The cell representing that block of time would only show the last measurements entered. The recommended timeline display would show each measurement and intervention. Rather than using a grid with standard-sized blocks, block size would be proportional to the time elapsed between actions. With this display, the user would be able to quickly see that several measurements were recorded and drugs administered within a short timeframe, and then activities became more regular as the patient was stabilized. At the beginning of the next shift, an incoming nurse could review the patient's flowsheet, see the activity, and immediately gain a better situational awareness about that particular patient's condition. Figure 4 shows a conceptual rendering of the timeline display.

By providing a visible indicator of elapsed time, the display draws attention to important events. It is also less likely that data will be hidden; even if recordings are too close together to show actual data values, the user can still see that values were recorded, an indicator that an event of interest occurred. It was further recommended that users could see data values via hover help , and that users could zoom in to a specific range of time to better see details.

The recommendation of a timeline view emerged from the main issue of the fixed-grid view hiding data. The usability team categorized this issue as a Type 1 decision-support issue, since hiding data from nurses and doctors could negatively impact a patients' health directly by hindering effective decision making. However, since implementing a timeline display would include reworking significant portions of the software, the team gave it a long term implementation timeframe. To immediately address the issue of meaningful data being hidden, the usability team recommended that the cells that contained hidden data should have color-coded triangles to indicate the nature of the obscured data. If a reading that has been hidden was out of a normal range, the triangle should be red, while all normal readings would warrant a green triangle. Because this was an easier thing to fix (the software already used triangles to show there were hidden data, and flagged out of range values with orange arrows, as seen in Figure 3), this recommendation was given an urgent implementation timeframe.

CONCLUSION

Because comprehensive EHR systems are just beginning to take hold in many non-federal healthcare facilities



End: 8/ 6/2010 @ 04:00 PM Resolution: 2 Hours

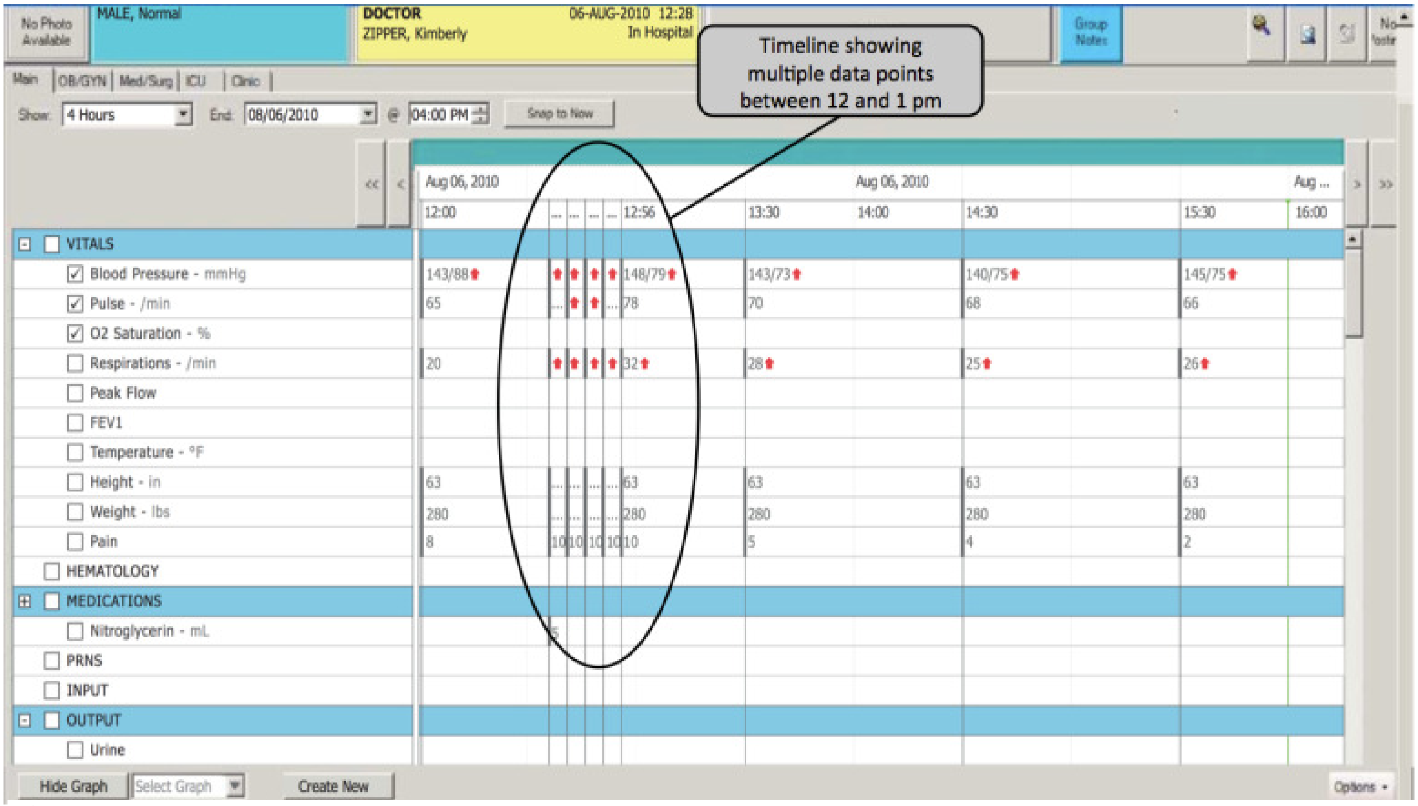
Triangle indicating multiple data points between 12 and 2 pm

Figure 3. Fixed grid showing cells with additional values

(only 12% non-federal hospitals have some form of an EHR, with less than 2% having a fully comprehensive system) (DesRoches et al., 2008; Jha et al., 2009), it is useful to look at the lessons learned from federal healthcare systems to lead the way in this technology. The Health Information Technology for Economic and Clinical Health (HITECH) Act brought increased attention to EHRs and we anticipate continued adoption and use of EHRs as a result. As more healthcare facilities work towards adopting and using EHRs, researchers and developers must take steps to ensure that usability guidelines are developed and implemented to allow for the successful integration of EHR technology into clinical settings. The current state of EHR dissemination provides the opportune environment to understand and develop such practices as described in this paper.

The use of a DCD approach in the context described here demonstrates the effective application of CSE and NDM concepts and methodology elements to a just-in-time usability analysis. We were able to apply modified versions of four out of the five phases of DCD within the constraints of the project. The first step of our rapid usability process, getting up to speed, is roughly equivalent to the preparation phase of DCD; we took time to familiarize ourselves with the domain, the user needs, and core workflow. The second step of the rapid usability process, establishing common ground, is akin to the second and third phases of DCD: knowledge elicitation and analysis and representation. During this step, we interviewed users, then analyzed their feedback to uncover areas that could benefit from cognitive support. The third step of the rapid usability process, developing recommendations, is similar to the application design phase of DCD. We presented design alternatives to the development teams; however, we did not actually build any of the new designs. Due to the constraints of the project, we were unable to do any formal evaluation (the fifth phase of DCD).

This approach could be particularly useful for future usability and design efforts that face similar constraints to the project described here but that could still benefit from the application of the DCD framework. Assigning issue classes and implementation timeframes to the issues and recommendations help the developers decide how to incorporate the findings from the usability analysis. This process accounts for multiple needs- the users' cognitive needs, along with the developers' needs to release software in a timely manner. Although not the traditional method of usability analysis, use of the DCD framework allowed the team to provide feedback with regard to critical design features and core functionality. Using cognitive support requirements as a framework from which to categorize issues and recommendations helps ensure that the EHR software will be able to



Timeline showing multiple data points between 12 and 1 pm

	12:00	...	...	...	...	12:56	13:30	14:30	15:30	16:00
☐ VITALS										
☑ Blood Pressure - mmHg	143/88↑	↑	↑	↑	↑	148/79↑	143/73↑	140/75↑	145/75↑	
☑ Pulse - /min	65	...	↑	↑	...	78	70	68	66	
☑ O2 Saturation - %										
☐ Respirations - /min	20	↑	↑	↑	↑	32↑	28↑	25↑	26↑	
☐ Peak Flow										
☐ FEV1										
☐ Temperature - °F										
☐ Height - in	63	...	...	...	...	63	63	63	63	
☐ Weight - lbs	280	...	...	...	...	280	280	280	280	
☐ Pain	8	10	10	10	10	10	5	4	2	
☐ HEMATOLOGY										
☐ MEDICATIONS										
☐ Nitroglycerin - mL		5								
☐ PRNS										
☐ INPUT										
☐ OUTPUT										
☐ Urine										

Figure 4. Timeline view showing an event, then stabilization

support healthcare workers' decision making and expertise in the complex environments in which they work.

AUTHOR NOTES

This activity was supported by a subcontract with the Indian Health Service (IHS) through Vangent, Inc. The opinions expressed herein are not those of the Indian Health Service, and no endorsement of the authors or Applied Decision Science, LLC, by IHS is expressed or implied. Correspondence regarding this paper should be directed to Christen Lopez [c.lopez@applieddecisionscience.com]

REFERENCES

Crandall, B., Klein, G., & Hoffman, R.R. (2006). Working minds: A practitioner's guide to cognitive task analysis. Cambridge, MA: MIT Press.

DesRoches, C. M., Campbell, E. G., Rao, S.R., Donelan, K., Ferris, T. G., Jha, A. K. ., . . . Blumenthal, D. (2008). Electronic health records in ambulatory care- a national survey of physicians. *New England Journal of Medicine, 359*(1), 50-60.

Grudin, J. (1994). Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM, 37*(1), 92-105.

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* (383-416). Mahwah, NJ: Lawrence Erlbaum Associates.

Jha, A. K., DesRoches, C. M., Campbell, E. G., Donelan, K., Rao, S. R., Ferris, T. G., . . . Blumenthal, D. (2009). Use of electronic health records in U.S. hospitals. *The New England Journal of Medicine, 360*(16), 1628-1638.

Militello, L. G., Dominguez, C. O., Lintern, G., Klein, G. (2009). The Role of Cognitive Systems Engineering in the Systems Design Process. *Systems Engineering Journal.*

Militello, L. G., Klein, G. (in press). Decision-Centered Design. *In J.D. Lee and A. Kirlik (eds.), The Oxford Handbook of Cognitive Engineering.* Oxford: Oxford University Press.

Norman, D.A. (1999). Affordance, conventions, and design. Interactions May-June, 38-44.

Shneiderman, B. & Plaisant, C. (2005). Designing the user interface: Strategies for effective human-computer interaction (4th ed). Boston: Pearson

Tools to Support Distributed Adaptive Planning: Airport Surface Displays for Collaborative Departure Management

Alicia Borgman Fernandes
Ohio State University

Philip J. Smith
The Ohio State University

As a distributed adaptive planning task, airport departure management provides a context for studying how shared displays can support a shift from resource-intensive collaboration and interaction to more efficient display-based coordination and synchronization. We report on observational studies of how these displays help distributed agents in a competitive-cooperative system: (a) Develop shared models of system state; (b) Anticipate and identify constraints faced by remote agents; (c) Observe actions of remote agents; (d) Adapt plans to variability in the world; (e) Reduce time and workload required to coordinate; and, (f) Improve timeliness of decision-making and problem solving. This has possible implications for the distribution of roles and responsibilities in the system. It also contributes to the knowledge required to customize information displays for agents in different roles and to refine models of distributed adaptive planning as a generic task.

KEYWORDS: Distributed work; adaptive planning; coordination; airport departure management; airport surface displays

INTRODUCTION

The National Airspace System (NAS) is a highly distributed, complex work system. Various agents in the system routinely perform adaptive planning tasks in situations with time-varying levels of uncertainty (Smith, Beatty, Spencer, & Billings, 2003; Smith et al., 1995). Managing the flow of departures from an airport is a prototypical task in the NAS that can require time-consuming, direct human-human collaboration to develop and adapt suitable air and surface traffic management plans.

Hayes-Roth and Hayes-Roth (1979), Suchman (1987) and others (Hollnagel & Woods, 2005; Zsambok & Klein, 1997) have described models of planning and adaptation. Key aspects of these models include an expert who recognizes some pattern in the state of the world and retrieves a strategy for managing a situation following that pattern. The expert modifies the retrieved strategy according to features of the current situation. Once the plan is implemented, the expert monitors the continued appropriateness of the plan as the situation in the world evolves and makes adjustments as necessary.

In addition to the adaptive planning cycle noted above, distributed adaptive planning requires coordination and collaboration at a distance—sometimes across competing organizations. This introduces additional challenges to the task of designing tools to support practitioners. For example, tools should align with the goals and responsibilities of each of the agents in the distributed work system (Grudin, 1988; Olson & Olson, 2000). In addition, tools should provide each agent with the information required to carry out their individual responsibilities (Smith, Spencer, & Billings, 2007), as well as allow them to observe the status and actions of others in the distributed system (Klein, Woods, Bradshaw, Hoffman, & Feltovich, 2004)—to the extent that organizations do not lose their competitive edge.

Airport departure management is an adaptive planning task that is distributed across organizations that have different goals and work in a competitive-cooperative environment (Smith et al., 1995). The dynamic nature of airport departure conditions requires continuous adaptation of management strategies to maintain safe and efficient traffic flows (Smith et al., 2003). These adaptations occur at multiple levels of abstraction (Hollnagel & Woods, 2005; Woods & Shattuck, 2000), such as shifts in overall strategy that affect several aircraft or changes impacting one or a few flights. The nature of the necessary adaptations depends on several factors, such as:

- The degree of uncertainty in conditions, such as uncertainty in how a weather system will develop (Smith et al., 2003).
- The length of the available planning horizon between the time that the need for adaptation is identified and when the change must be in place in order to have the desired effect (Zsambok & Klein, 1997).
- The amount of variability that can be tolerated and still maintain acceptable system performance (Hollnagel, Woods, & Leveson, 2006; Woods & Hollnagel, 2006).
- The lag time between the time a change is implemented by a human manager and the time the change is fully realized in the system (Hutchins, 1995).

Tools to support distributed adaptive planning in airport departure management should be informed by what is known about distributed work systems (Burke, Stagl, Salas, Pierce, & Kendall, 2006; Hinds & Kiesler, 2002; Olson & Olson, 2000; Smith, Spencer, & Billings, 2007; Woods & Branlat, 2010), as well as adaptive planning (Hayes-Roth & Hayes-Roth, 1979; Suchman, 1987; Zsambok & Klein, 1997). In addition, tools should support planning across varying levels of complexity, uncertainty, and time scales (Woods & Hollnagel, 2006). They should help people develop accurate models of the state of the system, anticipate how the system will change over the planning horizon, and provide appropriate feedback about how well the plan fits the changing situation (Smith, Bennett, & Stone, 2006; Smith, Stone, & Spencer, 2006; Woods & Hollnagel, 2006).

Airport surface displays represent a technology that can support distributed adaptive planning in airport departure management. They can be used as a shared display that can help the distributed team members develop a shared model of the underlying system, which supports distributed cooperative problem solving (Cannon-Bowers & Salas, 2001; Smith et al., 1995). In addition, as shared displays they can make visible the actions of remote actors and support distributed team members in anticipating the effects of those actions on the system (Hutchins, 1995; Klein et al., 2004; Olson & Olson, 2000). Airport surface displays provide users with feedback that can help them determine whether changes are needed in their current departure management strategies.

This paper is organized as follows: In the next section we describe airport departure management as a distributed adaptive planning task. This is followed by discussion of airport surface displays as an example of a tool to support distributed adaptive planning in airport departure management and observations of their use in the domain. We conclude with some thoughts about how further study of features of airport surface displays supporting airport departure management can improve designs for different roles in this domain as well as inform models of distributed adaptive planning as a generic task.

DISTRIBUTED ADAPTIVE PLANNING IN DEPARTURE MANAGEMENT

Airport departure management requires planning at multiple levels and across varying time scales. At a strategic level, traffic managers plan traffic flows in anticipation of ground and airspace constraints such as weather systems. Similarly, Air Traffic Control Tower (ATCT) personnel develop strategies for queuing aircraft for departure according to runway configuration and departure route availability. Such strategies must be adapted as conditions change over time. Conditions impacting departures can change with varying prediction horizons and levels of uncertainty. For example, some weather systems can be highly unpredictable, making it difficult to anticipate whether certain departure routes will be available within the next hour. Weather systems are a main cause of delays in aviation.

Traffic management personnel and flight operator Air Traffic Control (ATC) Coordinators have reported that when provided a weather forecast, they mentally develop a scenario indicating how they expect the weather system to impact air traffic in their area. They use previous experience with similar weather systems to select from a set of high-level strategies for coping with the degree of uncertainty and forecast lead time typical of similar weather systems. The most appropriate high-level strategy is then adapted to fit their expectations for the weather system currently under consideration.

Using this information, traffic managers and ATC coordinators confer via teleconference to share their expectations for the impact of the weather and preferred strategies for coping with the weather impacts on air traffic. The teleconference is used to communicate the strategy to others, although traffic managers may refine their strategies based on input from others in the teleconference.

Throughout the weather event, traffic management personnel monitor the actual development of the weather system and evaluate whether their expectations were correct. If departure conditions do not meet their expectations, such as if the weather system develops differently than predicted, traffic management personnel also must evaluate whether their current strategy is adequate to cope with the situation, or if it should be modified.

For example, flight operator personnel interviewed at one airport reported that they abstractly characterize most thunderstorms as either "frontal" or "popcorn" storms. Both disrupt air traffic, but there are different strategies for coping with them. "Frontal storms are predictable… the dispatcher can get around the weather by planning for it." That is, such storms have a fairly low level of uncertainty about which routes will be unavailable to air traffic as well as a fairly high amount of lead time on the prediction. The flight operator employs a strategy for planning routes based on these reliable predictions about what routes will be available when. On the other hand, popcorn storms are characterized by a high level of uncertainty about when a given route will be unavailable for air traffic as well as a short time between a prediction that a given route will be impacted by the weather and the time that the route closes. The flight operator must employ a different strategy for planning routes under such conditions.

When the weather exhibits a higher level of uncertainty and/or a shorter time horizon for reliable predictions than the frontal storms described above, flight operators are not always able to file routes that will be available when the flight actually departs. In such cases, Air Route Traffic Control Center (ARTCC) and ATCT traffic management personnel coordinate pre-departure route amendments for flights whose routes become unavailable. The typical process for processing pre-departure route amendments is described below.

Pre-Departure Route Amendment Process

Often, when weather conditions exhibit high levels of uncertainty and/or short forecast lead times, multiple departures taxiing out have filed routes that are unavailable. In such cases, ATCT personnel provide ARTCC traffic managers a list of flights whose routes are (or are expected to be) unavailable. The ARTCC then selects an alternate route for each flight and confirms the set of route amendments with the ATCT. ATCT controllers read the amended route to the flight crew. The flight crew then confirms the change with the flight dispatcher to ensure that fuel, equipment, and other requirements are met for the new route. Note that this process is typical across the United States, although there are exceptions. Some facilities can send the amended clearance to the flight crew digitally and the ATCT controller does not have to read the entire clearance to the flight crew. Some facilities and flight operators also pre-coordinate acceptable route amendments and so flight crews do not have to confirm route amendments with their dispatchers.

The typical pre-departure route amendment process, however, can break down because of delays that preclude a timely response. First, it can require multiple telephone calls, particularly when conditions are dynamic and departure fix availability is uncertain. These telephone calls can be time-consuming, particularly since the situation can require coordination among ARTCC Traffic Managers, ATCT Traffic Management Coordinators, and Flight Operator ATC Coordinators. This process must be repeated for each flight requiring a route amendment.

In addition, flights may not receive their pre-departure route amendments in the order in which aircraft will reach the runway threshold for departure. Although it would be more efficient for route amendments to be processed in the order in which flights will be ready to depart, this order is not always visible to the ARTCC (typically ARTCC traffic managers do not have a view of the airport surface). If a departure arrives to the runway threshold without an available route, a number of departures may be subject to additional delay.

To support these complexities, traffic management personnel have access to a variety of tools such as weather forecasts that facilitate anticipation of future route availability. In addition, they can collaborate via telephone and radio communications. However, typically only ATCT personnel have a view of the entire airport surface. Providing ARTCC, TRACON, and flight operator personnel a remote view of the airport surface can provide opportunities for coordination and information sharing that support distant agents in developing a common model of system status. Thus, airport surface displays can complement other tools that support distributed adaptive planning in airport departure management.

AIRPORT SURFACE DISPLAYS IN DEPARTURE MANAGEMENT

A view of the airport surface can provide key information useful for planning and decision-making. At most airports, only personnel located in the Air Traffic Control Tower (ATCT) have a full view of the airport surface. At times, traffic managers at other facilities can benefit from such information for their decision-making. Without a view of the airport surface, they obtain such information via a telephone call to the ATCT. While decision making at the TRACON and ARTCC level can benefit from input from local airport personnel, collaboration and information exchange via telephone can be time-consuming. In addition, the timeliness and accuracy of telephone information exchange can be affected by the workload of personnel in the ATCT, ARTCC, and other facilities.

Flight operator personnel also can benefit from information about aircraft locations and delays on the airport surface. Without a surface display, they depend on information from the ATCT, ramp control personnel, or individual flight crews. Exchange of this information occurs via telephone or radio and also can depend on the workload of the personnel located on the airport surface. Although ramp control personnel often are located in an elevated Ramp Control Tower (RCT), these towers do not always provide visibility of the entire airport surface. Flight crews also have a limited view of the surface and therefore may be limited in the information they can provide beyond the status of their aircraft (e.g., its location in the departure queue).

Airport surface displays are being installed at an increasing number of NAS facilities. These provide a map display of the airport surface, including locations of all aircraft with an active transponder. They also can provide statistics generated from the location data. Such information can enable new and powerful forms of distributed coordination and collaboration that support NextGen Collaborative Air Traffic Management (CATM) goals (Federal Aviation Administration, 2010). The nature of surface displays also supports traffic managers in identifying how their plans should be adapted to account for changes in the system state. In airport departure management, such changes typically involve changes in the capacity of one or more departure routes such as due to changes in weather.

Previous Surface Display Research

Research into the utility of surface displays for supporting distributed adaptive planning exists, although most of the findings seem to highlight the decision support capabilities included with the surface display rather than the utility of the surface display itself. For example, the Surface Management System (Atkins, Brinton, & Walton, 2002; Spencer, et al., 2003; Spencer, Smith, Billings, Brinton, & Atkins, 2002; Walton, Quinn, & Atkins, 2002) included a surface display as well as tools to support decision-making and information sharing regarding surface management issues such as runway configuration changes, runway balancing, and strategies for building departure queues (Spencer et al., 2002; Spencer et al., 2003).

Spencer et al. (2002) discuss support for such surface management tasks from the point of view of an individual ATCT Traffic Management Coordinator (TMC). However, there also have been studies involving the ability to share information to support distributed planning. For example, Spencer, Smith, and Billings (2005) discuss the use of such systems for sharing context-sensitive information and alerts regarding aircraft status such as late arrivals, spot conflicts, flights in danger of missing the departure window assigned by a Ground Delay Program, runway assignment changes, and route status. Such views also can aid RCT and ATCT personnel in anticipating and preventing conflicts at gate alley and taxiway choke points (Spencer, Smith, & Billings, 2001).

A surface display makes available information about aircraft location and status to traffic managers not located in the ATCT as well. In particular, remote traffic managers can use the surface display to identify flights that have pushed back from the gate but have not yet entered the active movement surface. Currently, most traffic managers hesitate to consider in their decision-making aircraft that are not yet taxiing under ATCT control because the time at which a flight will actually be ready to taxi out for departure can be difficult to predict with certainty. A view of aircraft taxiing in the ramp area can provide earlier, high quality information about the set of flights that will soon be able to depart, giving traffic management personnel more time to adjust departure management strategies to fit current weather and traffic conditions.

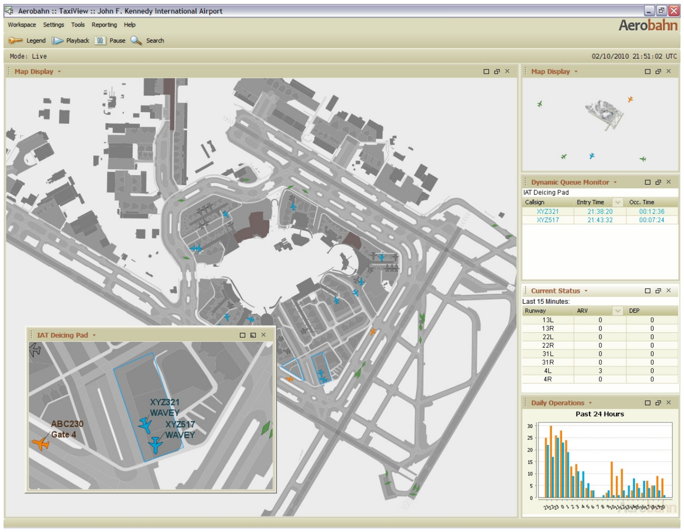
Traffic management and flight operator personnel can use aircraft location as well as information included in the data tag associated with each aircraft icon in their departure management decision making. Data tags provide information such as call sign, destination, and departure fix that is not readily available from viewing the physical aircraft on the surface – if the person can see the aircraft at all – and can provide clues about how best to manage the aircraft on the airport surface.

Surface displays have been installed in a number of NAS facilities and are used in a variety of ways. Interview descriptions and observations of uses of such displays are discussed in the next section.

Observed Surface Display Uses

As part of a larger study of the integrated management of airspace and surface constraints, observational studies were carried out at several NAS facilities. The observations described here came from the ATCTs of two major US airports, two ARTCCs, and two RCTs for hub operators. Personnel responsible for traffic management and coordination tasks were interviewed about and observed in their use of the displays. Observations and interviews were made over the course of two weeks during the summer of 2010.

The ATCTs visited were equipped with the Airport Surface Detection Equipment, Model X (ASDE-X). The ARTCCs and RCTs were equipped with Aerobahn. An example image of Aerobahn is shown in Figure 1. The key feature of such displays is a map of the airport, including runways, taxiways, and buildings. The map is populated with icons indicating the location (and typically identification) of each aircraft on the surface. The map also can indicate locations of ground vehicles on the active surface and in ramp areas. Several systems include other displays such as aircraft that are airborne in the vicinity of the airport, shown in the upper right corner of Figure 1.



Dynamic Queue Monitor

IAT Deicing Pad

Callsign	Entry Time	Occ. Time
XYZ321	21:38:20	00:12:36
XYZ517	21:43:32	00:07:24

Current Status

Last 15 Minutes:

Runway	ARV	DEP
13L	0	0
13R	0	0
22L	0	0
22R	0	0
31L	0	0
31R	0	0
4L	3	0
4R	0	0

Figure 1. Sample image of Aerobahn showing a map of the airport surface, aircraft locations and identifications, aircraft in nearby airspace, and descriptive statistics. Used with permission from the Sensis Corporation.

The displays were observed to support several different tasks. In most cases the surface display supported development of shared models that improved coordination in distributed adaptive planning. However, the displays also were used to facilitate individuals' decision-making. Some of the key observed uses are described here. Specifically, remote agents used them to monitor airport surface and airport departure procedure status. Displays were customized to alert users to aircraft that required their attention and such customizations could directly support coordination processes. However, they also exhibited some limitations.

Remote Awareness of Airport Surface Status

Personnel without a direct view of the airport surface as from an ATCT were observed to benefit directly from the digital view of the locations of aircraft on the airport surface. The airport surface display integrates data from physical sensors in a way that facilitates building mental representations of system status when that system cannot be directly viewed (Bowers, Salas, & Jentsch, 2006; Endsley, Bolte, & Jones, 2003; Hollnagel & Woods, 2005; Hutchins, 1995; Olson & Olson, 2000; Smith, Bennett, & Stone, 2006). ARTCC traffic managers do not have a direct view of the airport surface, but their traffic management decisions often are influenced by the traffic situation at the airport. Typically, the traffic managers' sensing agents are ATCT traffic management personnel and the means of communicating surface status information is the telephone. Such users of surface displays reported that the display provided much improved awareness of the status of flights on the airport surface. One ARTCC Traffic Manager said, "We've never had that before."

This remote awareness of the status of flights on the airport surface improves the ability of ARTCC traffic managers to anticipate the future status of traffic demand for parts of their airspace. Tools that support anticipation invoke traffic managers' expertise (Smith, Bennett, & Stone, 2006; Woods & Hollnagel, 2006), helping them to develop more adaptive strategies for managing traffic flows and, in turn, controllers' workload. Specifically, the surface display provides information allowing traffic managers to make earlier, more precise predictions about the time at which to expect specific flights—before the flights leave the ground as opposed to after they are airborne.

Without the specific information about the actual time-varying demand for specific departure routes that the surface display provides, traffic managers invoke traffic management strategies such as miles in trail to ensure sufficient spacing between aircraft when they expect demand for a given route to be high. Such approaches can decrease demand more than is necessary to manage controller workload. This also can cause available capacity to go unused, decreasing efficiency. Alternatively, the surface display allows ARTCC traffic managers to view the actual demand for the airspace and invoke a more efficient strategy to ensure that controllers do not experience too much traffic at once: "I can look at [the surface display] and see where [departures to that route] are and ask the guys in the Ramp Tower [RCT] to space them out a little bit coming out" (ARTCC traffic manager). The traffic manager can then use the surface display to monitor the actions of the RCT and ATCT and very quickly evaluate whether they create sufficient space between flights to manage demand and controller workload. That is, the surface display provides a mechanism for very quickly gaining feedback about whether the actions are likely to be sufficient (Smith, Stone, & Spencer, 2006).

This ability to better anticipate airspace demand also can enable ARTCC traffic managers to make faster, more informed decisions about when to implement strategies for coping with impacts of weather: “If we will need to close [a departure route] soon, currently we have to call the Tower [ATCT] to ask how many flights are in the queue to [that route] and work out how many more flights to send through [that route] before they start rerouting flights. With [the surface display] we can look at the display and see how many flights are in the queue for [that route] and then call the Tower [ATCT] to tell them that … flight needs a reroute” (ARTCC traffic manager).

Timeliness of decision-making also can be determined by the order in which a series of problems is solved. As noted above, the current telephone-driven process by which the ATCT provides a list of flights to the ARTCC that need pre-departure route amendments does not guarantee that the ARTCC processes route amendments for flights in the order in which the flights will take off. Specific knowledge about the locations of aircraft on the airport surface enables ARTCC traffic managers to ensure they provide the first route amendment to the departure closest to the runway threshold. Using the surface display in the ARTCC “gives them [the ATCT and RCT] the reroutes in the right order,” according to one ARTCC traffic manager. Without the surface display, the ARTCC traffic manager said, “Sometimes we could get the number one at the runway” when selecting the first flight to reroute, but not always.

Monitoring Metering Program Status

ARTCC traffic managers are not the only remote agents charged with airport departure management decision-making that were observed to benefit from the use of a surface display. At one airport where a departure metering program is in place, program managers do not have a direct view of the airport surface. Instead, they were observed to use a surface display to gain direct feedback about the appropriateness of their metering plan in the current departure scenario (Fernandes & Smith, 2011; Fernandes et al., 2010). It also provided them information about how they should modify program parameters to account for changes in the situation. In particular, they were observed to use the surface display to track the number of aircraft in the final departure queue, a key metering program parameter.

Customized Data Blocks

In addition to providing remote viewers key information to improve their decision making, surface displays can be used to improve coordination and collaboration processes and help users manage their attention. For example, organizations and individuals were observed to have customized aircraft icons and data blocks on the surface displays, most commonly through the use of color. The most common purpose of these customizations was to add context-sensitive alerts that would focus users’ attention on aircraft that should be a high priority for a variety of reasons.

At all facilities where users of the Aerobahn surface display were observed, aircraft icons and data blocks were color-coded to indicate aircraft status. At both of the RCTs visited, a departure with a crew that would time out if the flight was not airborne soon would be denoted by a color and message indicating its status as in danger of a crew time-out. The data block information, coupled with the aircraft location on the surface, aids RCT personnel in reasoning about the likelihood that the flight will be able to depart before the crew times out. They may be able to coordinate with the ATCT to expedite its departure, or they may need to coordinate with the ATCT to return the flight to its gate.

Similarly, aircraft data blocks also were configured to alert RCT personnel when an aircraft became a high priority while it was taxiing out. For example, facilities were observed to have their surface displays configured to identify departures according to the length of time they had been Out (i.e., pushed back from the gate). At certain thresholds with respect to the Department of Transportation (DOT) tarmac delay regulation (Department of Transportation, 2009) the icon and/or data block could change to elevate that flight’s priority. According to multiple RCT Managers, alerts on the surface display indicating the amount of time a departure has been Out cause RCT personnel to “see if there’s a way to get him Off [airborne]” or “start working on getting him back to the gate” to ensure that passengers would have the opportunity to deplane before the three hour limit.

These context-sensitive alerts, coupled with a remote view of the airport surface also can reduce the need for time-consuming, direct human-human collaboration (typically telephone calls). One RCT Manager reported that tools such as the crew timeout alerts "have cut down on phone calls… Flight Ops [the Flight Operations Center (FOC)] used to call us to ask how long he'd [a specific flight] taxi [to get to the runway] because he'll time out." The ability of the RCT to see which flights are in danger of timing out and for the FOC to see how close that aircraft is to the runway allows them to limit their communication and collaboration with each other to those cases where the RCT believes that the aircraft will not get to the runway before the crew times out and must return to the gate. This saves time for both parties, reducing their workload.

Coordination via Surface Display

Facilities in at least one ARTCC have taken the customized data blocks even further. Traffic management personnel in the ARTCC and the ATCT, as well as the RCT operator at the main airport within the jurisdiction of the ARTCC, use customized data blocks to quickly coordinate pre-departure route amendments. Thus, they use the surface display as a shared display to mediate coordination through asynchronous communication.

In particular, according to an RCT manager, the process is "used when a departure gate suddenly closes and there are a lot of departures". Personnel change the color of aircraft icons and add messages to aircraft data blocks to alert others to the status of individual departures. The RCT can request a route amendment for a departure, the ARTCC can indicate that the route has been amended, and the ATCT can indicate that the route amendment has been received and the amended clearance delivered to the flight crew. An RCT manager provided the following example:

"If West flights need to be rerouted, the Center [ARTCC] will call the Ramp Tower [RCT] and say, 'Starting with Flight 220 (for example) and after, everyone needs a reroute.'" RCT personnel select the aircraft icon on the surface display and add an alert that changes the icon on the RCT display and on the surface display in the ARTCC. The action changes the color of the aircraft icon and adds a message to its data block stating "FAA—Need Reroute."

"The Center [ARTCC] TMC sees the alert, puts the reroute in, and changes" the data block and alert on the surface display. The changes are propagated to the RCT and ATCT. The aircraft icon now has an "FAA—Clearance" message in its data block. "Now [the flight is] waiting to get the clearance from the Tower [ATCT]. Once we get it we [the RCT] change the tag to 'Rerouted'."

This pre-departure amendment process, according to one ARTCC traffic manager, "cuts down on the communication. They [the ATCT and/or RCT] don't have to call us with five or six aircraft, wait two minutes and call again with five or six more." In addition, "it cuts down on the miscommunications."

This procedure is faster than the telephone process described above, making it particularly useful for highly dynamic situations in which there is sufficient uncertainty in the weather that it is difficult to anticipate which departure routes will be available twenty to thirty minutes into the future. In addition, this process represents effective exploitation of some of the advantages of the surface display for traffic management personnel remote from the airport.

In particular, the process represents use of a shared digital display for coordination (Bowers, Salas, & Jentsch, 2006; Endsley, Bolte, & Jones, 2003; Hinds & Kiesler, 2002; Olson & Olson, 2000) between the ARTCC, ATCT, and RCT. The shared display is useful in this setting because its content is relevant to all parties. As such, it helps the distributed team develop a common representation of the traffic situation when conditions pose high levels of uncertainty. Each of the agents is able to accomplish tasks under their individual responsibility, but using the surface display they can coordinate more effectively with each other (Smith, Spencer, & Billings, 2007; Woods & Shattuck, 2000). All of the relevant personnel interviewed for this study reported that the process reduces the workload involved in coordinating pre-departure route amendments and departure fix spacing needs.

Even without a formal procedure for coordinating pre-departure route amendments, the surface display can facilitate such coordination. For example, an RCT Manager at a different airport (under the jurisdiction of a different ARTCC) reported that during dynamic weather conditions, delayed aircraft are typically parked on taxiways near the runway. "We use [the surface display] to answer the question: can you get him to the runway? They [the ARTCC] should be able to look at their [surface] display and see if there's a route to the runway" when coordinating a reroute.

Thus, surface displays help address issues in the distribution of roles and responsibilities relative to access to important data and knowledge in the NAS (Hinds &

Kiesler, 2002; Smith, Spencer, & Billings, 2007). Typically, the ARTCC is responsible for processing route amendments but does not always have the best information about which departures need route amendments and the best order in which to process them. The surface display can provide this information.

Surface Display Limitations

Although surface display tools provide useful (and usually accurate) information to several organizations and support coordination, they are not a panacea. As with any technology, they show signs of brittleness (Roth, Bennett, & Woods, 1987; Smith, McCoy, & Layton, 1997). Although infrequent, observations at multiple facilities identified missing aircraft, targets with incomplete or inaccurate data blocks, and even the system going down altogether. The surface display technology relies on sensors located at various locations on the airport surface, transponders located on each aircraft, and data transmission systems, any of which can fail. “The transponder has to work,” said one RCT Manager. It also requires the flight crew to turn the transponder on when the aircraft is ready to push back. Otherwise, the aircraft either may not have an icon at all, or it may appear as an unidentified object.

In addition, there are times when the surface display incorrectly identifies an aircraft, which can cause breakdowns in shared mental models. For example, during observation at one RCT, a departure was given a new route and started taxiing toward the runway. There was some confusion about where the flight was because the Ramp Controller could not find it on the surface display. The RCT Manager knew which flight was taxiing out and where it was on the surface (i.e., he could identify it by looking out the RCT window) but the surface display assigned a different call sign to that flight. (“He was given a [South] route but the [data block] changed.”) It took some time for the Manager to convince the Ramp Controller that he was right about the actual location of the aircraft in question and that the surface display was wrong.

As with any technology, operators require adequate training to most effectively use the surface display systems (Olson & Olson, 2000). However, many personnel with access to surface displays may not have received adequate training to make full use of their capabilities. For example, two different Managers at one RCT were interviewed: one that typically works during the morning and one that typically works during the (busier) evening hours. The daytime Manager had discovered many more features of the surface display system than the evening Manager because he had spent more time learning to use the system during his shift. It is not clear whether this is because of differences in workload between the two shifts or personal differences in the two Managers. Nevertheless, it serves as a reminder that operators need to receive adequate training on the systems provided to support their work in order to effectively exploit the capabilities of the system. Regardless of the amount of training they receive on the features of a system, operators will choose to use the technology only to the extent that they can see its benefit to their work (Grudin, 1988; Hutchins, 1995; Olson & Olson, 2000).

CONCLUSION

Surface displays represent a key emerging technology that can facilitate distributed adaptive planning as well as more efficient coordination and collaboration in airport departure management. The display is a representation of a key part of the NAS that allows distant practitioners to directly perceive anticipated bottlenecks and makes salient potential solutions and constraints on solutions (Smith, Bennett, & Stone, 2006; Smith, Stone, & Spencer, 2006; Vicente & Rasmussen, 1990). As such, they also make visible the actions taken by distant agents (Hutchins, 1995; Olson & Olson, 2000) and make salient important impacts of these actions or potential impacts of these actions (Smith, Stone, & Spencer, 2006).

Visibility of the actions of distant agents supports coordination in part by reducing the amount of time-consuming communication required to develop a shared understanding of system status and to identify needs and constraints faced by distant agents (Cannon-Bowers & Salas, 2001; Klein et al., 2004; Olson & Olson, 2000; Smith et al., 1995). Instead, airport surface displays enable coordination based on the contents of the display itself. They support asynchronous coordination to relieve bottlenecks as well as synchronous collaboration to adapt strategies to changing conditions. Such improvements in distributed coordination, collaboration, planning and adaptation can improve the efficiency of the NAS without compromising safety.

Note that the airport surface display as a representation of domain constraints (Smith, Bennett, & Stone, 2006; Vicente & Rasmussen, 1990) is incomplete. Further work is needed to customize the representations provided to different people to support their individual roles and responsibilities in airport surface management. This work needs to consider the ability of the display to change the nature of coordination in the domain, potentially changing the current distribution of roles and responsibilities in the NAS (Olson & Olson, 2000; Smith et al., 2003; Smith et al., 1 995; Smith, Spencer, & Billings, 2007; Woods & Hollnagel, 2006).

In order to effectively customize surface display representations for different people in the NAS, it also will be important to understand what specific features of the display personnel in different roles use to achieve an accurate model of system state, identify constraints, and develop and adapt plans accordingly (Endsley, Bolte, & Jones, 2003; Smith, Bennett, & Stone, 2006; Smith, Stone, & Spencer, 2006; Vicente & Rasmussen, 1990; Woods & Hollnagel, 2006). Such information can be used to inform designs that enhance direct perception and knowledge-based problem solving (Smith, Bennett, & Stone, 2006; Vicente & Rasmussen, 1990). It also can inform models of adaptive planning as required by each role, distributed adaptive planning in the domain, and distributed adaptive planning as a generic task. This approach to using increasingly abstract characterizations can inform specific designs as well as contribute to general knowledge about joint cognitive work (Woods & Hollnagel, 2006).

AUTHOR NOTES

The FAA Human Factors Research & Engineering Group coordinated the research requirement and its principal representative acquired, funded, and technically managed execution of the research service. In addition, the authors wish to thank the individuals at the Federal Aviation Administration (FAA), airport operator, and flight operators that made the visits to facilities possible. The authors wish to thank the many personnel who granted interviews for the study. Their insights are invaluable. Correspondence regarding this paper should be directed to Alicia Borgman Fernandes [adborgman@gmail.com]

REFERENCES

Atkins, S., Brinton, C., & Walton, D. (2002). Functionalities, displays, and concept of use for the surface management system. *21st AIAA/IEEE Digital Avionics Systems Conference.* Irvine, CA.

Bowers, C., Salas, E., & Jentsch, F. (Eds.). (2006). *Creating high-tech teams: Practical guidance on work performance.* Washington, DC: American Psychological Association.

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: A conceptual analysis and model. *Journal of Applied Psychology, 91*(6), 1189-1207.

Cannon-Bowers, J. A., & Salas, E. (2001). Reflections on shared cognition. *Journal of Organizational Behavior, 22*, 195-202.

Department of Transportation. (2009). Enhancing airline passenger protections. *Federal Register, 74249*, 68983-69004.

Endsley, M. R., Bolte, B., & Jones, D. G. (2003). *Designing for situation awareness.* London: Taylor and Francis.

Federal Aviation Administration. (2010, May 5). *NextGen Solution Set Collaborative Air Traffic Management (CATM).* Retrieved December 22, 2010, from http://www.faa.gov/about/initiatives/nextgen/portfolio/sol_sets/catm/index.cfm

Fernandes, A. B., & Smith, P. J. (2011). Human-centered design to support flexibility and adaptability in airport surface management. *16th International Symposium on Aviation Psychology.* Dayton, OH.

Fernandes, A. B., Smith, P. J., Durham, K., Evans, M., Beatty, R., & Spencer, A. (2010). *Airport Surface Management in Response to a Terminal Radar Outage.* Ohio State University. Columbus, OH. December 2010.: Cognitive Systems Engineering Laboratory.

Grudin, J. (1988). Why CSCW applications fail: Problems in the design and evaluation of organizational interfaces. *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work.* New York: ACM.

Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science, 3*(4), 275-310.

Hinds, P., & Kiesler, S. (Eds.). (2002). *Distributed work.* Cambridge, MA: MIT Press.

Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations in cognitive systems engineering.* Boca Raton, FL: CRC Press.

Hollnagel, E., Woods, D. D., & Leveson, N. (2006). *Resilience engineering: Concepts and precepts.* Adelshot, UK: Ashgate.

Hutchins, E. (1995). *Cognition in the Wild.* Cambridge, MA: MIT Press.

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems, 19*(6), 91-95.

Olson, G. M., & Olson, J. S. (2000). Distance matters. *Human-Computer Interaction, 15*(2/3), 139-178.

Roth, E. M., Bennett, K. B., & Woods, D. D. (1987). Human interaction with an "intelligent" machine. *International Journal of Man-Machine Studies, 27*, 479-525.

Smith, P. J., Beatty, R., Spencer, A., & Billings, C. E. (2003). Dealing with the challenges of distributed planning in a stochastic environment: Coordinated contingency planning. *Proceedings of the 2003 Annual Conference on Digital Avionics Systems.* Chicago, IL. October 12-16, 2003.

Smith, P. J., Bennett, K. B., & Stone, R. B. (2006). Representation aiding to support performance on problem-solving tasks. In R. Williges (Ed.), *Reviews of Human Factors and Ergonomics* (Vol. 2, pp. 49-73). Santa Monica, CA: Human Factors and Ergonomics Society.

Smith, P. J., McCoy, C. E., & Layton, C. (1997). Brittleness in the design of cooperative problem-solving systems: the effects on user performance. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, 27*(3), 360-371.

Smith, P. J., McCoy, C. E., Orasanu, J., Billings, C. E., Denning, R., Rodvold, M., Gee, T. (1995). Cooperative problem-solving activities in flight planning and constraints for commercial aircraft. *Proceedings of the 1995 Annual Meeting of the IEEE Society for Systems, Man, and Cybernetics*, (pp. 4563-4568). Vancouver, British Columbia, Canada. October 23-27, 1995.

Smith, P. J., Spencer, A. L., & Billings, C. E. (2007). Strategies for designing distributed systems: Case studies in the design of an air traffic management system. *Cognition, Technology, and Work, 9*, 39-49.

Smith, P. J., Stone, R. B., & Spencer, A. (2006). Design as a prediction task: Applying cognitive psychology to system development. In W. Marras, & W. Karwowski (Eds.), *Handbook of Industrial Ergonomics* (2nd ed., pp. 24-1 - 24-18). New York: Marcel Dekker.

Spencer, A., Smith, P. J., & Billings, C. E. (2001). *Sample scenarios provided by an airport coordination center at JFK.* Technical Report #2001-18, The Ohio State University, Institute for Ergonomics, Columbus, OH.

Spencer, A., Smith, P. J., & Billings, C. E. (2005). Airport resource management and decision aids for airlines. *Proceedings of the 2005 International Symposium on Aviation Psychology.* Oklahoma City, OK.

Spencer, A., Smith, P. J., Billings, C. E., Brinton, C., & Atkins, S. (2002). Support of traffic management coordinators in an airport air traffic control tower using a surface management system. In S. Chatty, J. Hansman, & G. Boy (Ed.), *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics* (pp. 105-111). Cambridge, MA: AAAI Press.

Spencer, A., Smith, P. J., Billings, C. E., Brinton, C., Atkins, S., & Walton, D. (2003). Decision support and information exchange to improve airport surface management. *Proceedings of the 12th International Symposium on Aviation Psychology.* Dayton, OH.

Suchman, L. (1987). *Plans and situated actions: The problem of human-machine communication.* Cambridge, England: Cambridge University Press.

Vicente, K. J., & Rasmussen, J. (1990). The ecology of human-machine systems II: Mediating "direct perception" in complex work domains. *Ecological Psychology, 2*(3), 207-249.

Walton, D., Quinn, C., & Atkins, S. (2002). Human factors lessons learned from a surface management system. *AIAA' Aircraft Technology, Integration, and Operations (ATIO) Technical Forum.* Los Angeles, CA.

Woods, D. D., & Branlat, M. (2010). Hollnagel's test: Being "in control" of highly interdependent multi-layered networked systems. *Cognition, Technology, and Work, 12*(2), 95-101.

Woods, D. D., & Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering.* Boca Raton, FL: CRC Press.

Woods, D. D., & Shattuck, L. G. (2000). Distant supervision--local action given the potential for surprise. *Cognition, Technology, and Work, 2*, 86-96.

Zsambok, C., & Klein, G. (Eds.). (1997). *Naturalistic decision making.* Mahwah, NJ: Erlbaum Associates.

Notes for Contributors to *Cognitive Technology*

Manuscript Submission – Manuscripts may be submitted electronically directly to the editor (Stephen Fiore, sfiore@ist.ucf.edu) or the managing editor, (Brad Brubaker, brad.brubaker@indstate.edu), or to any associate editor (see inside front cover).

Manuscript style – All submissions must include a title, a short title for the running head, an abstract (less than 150 words), the names and affiliations of all authors, a list of 4 key terms (to aid electronic searches), a reference section and complete contact information (including address, phone number, and email address) of the principal author.

Format – All submissions should be written in English and attached to an e-mail as a Word document.

References – A list of all reference material should be included in APA (American Psychological Association) style. The formatting instructions contained in the 5th edition published in 2005 is acceptable. Internet references need to include the complete URL address and the date of access.

Figures and Illustrations – No figures, tables or illustrations should be embedded in the text. Instead include all illustrations as separate files or at the end of the reference section.

Procedure – All submissions are examined by at least two reviewers. Submitted manuscripts are e-mailed to reviewers to speed the reviewing process. Principal authors should receive feedback 4 to 6 weeks after submission.

Website Information – Additional submission information can be found at the Cognitive Technology Website at:
http://www.cognitivetechnologyjournal.com

The Society for Applied Research in Memory and Cognition (SARMAC)

SARMAC, established in 1994, provides activities and services for member researchers concerned with the application of research into memory and cognition.

Research Meetings

1995 SARMAC I Vancouver, BC, Canada
1996 Symposium on Memory Improvement and Rehabilitation at IMC in Padua, Italy
1997 SARMAC II Toronto, ON, Canada
1999 SARMAC III Boulder, CO, USA
2000 MiniConference at APS, Miami, FL, USA
2001 SARMAC IV Kingston, ON, Canada
2003 SARMAC V Aberdeen, Scotland
2005 SARMAC VI Wellington, New Zealand
2005 Second International Conference on Prospective Memory 2005 in Zurich, Switzerland
2006 The Fourth International Conference on Memory (ICOM) Sydney, Australia
2007 SARMAC VII Lewiston, ME, USA
2009 SARMAC VIII Kyoto, Japan
2009 Decade of the Mind IV Conference - Reverse Engineering the Brain Santa Ana Pueblo, NM, USA
2011 SARMAC IX New York, NY, USA
2013 SARMAC X Rotterdam, the Netherlands

PUBLICATIONS

The **Journal of Applied Research in Memory and Cognition** (JARMAC) published by Elsevier. JARMAC will feature innovative, creative empirical research targeting the overlap between cognitive theory and real-world application.

The SARMAC Newsletter is published twice a year, is delivered electronically and includes professional and academic information and notices of upcoming conferences in Memory and Cogniton.

The International Journal *Cognitive Technology*® is the official journal of the Practical Memory Institute and is cosponsored by SARMAC and published by Compact Disc Inc. Electronic access to current issues at http://cognitivetechnologyjournal.com and access to all past issues comes with SARMAC membership.

To become a member or renew membership, please visit www.sarmac.org; and click on the "membership" navigation box, where you will find a link to the current application form. Questions about membership and/or the society in general should be directed to the Executive Director, Michael Toglia, University of North Florida (m.toglia@unf.edu).

THE INTERNATIONAL JOURNAL

cognitive technology®

Human Factors

Cognitive Rehabilitation

Modeling and Simulation

Cognitive Systems Engineering

Cognitive Models and Agent Technologies

Since 1996

The promise of
cognitive psychology
lies in the progress of
cognitive technology