CH XX: Human Supervisory Control Challenges in Network Centric Operations

M.L. Cummings
P.M. Mitchell
T.B. Sheridan

Massachusetts Institute of Technology
Cambridge, MA

ABSTRACT

Network centric warfare (NCW) is a concept of operations that seeks to increase combat power by linking battlespace entities to effectively leverage information superiority. A network centric force must be supported by sophisticated automated systems, so human-computer interactions are an important aspect of overall performance. The Department of Defense (DoD) has recognized that a lack of understanding of human decision making relevant to NCW is a significant barrier limiting NCW's potential benefits. To this end, this report identifies ten human supervisory control challenges that could significantly impact operator performance in NCW: Information overload, appropriate levels of automation, adaptive automation, distributed decision-making through team coordination, complexity measures, decision biases, attention allocation, supervisory monitoring of operators, trust and reliability, and accountability. Network-centric operations will bring increases in the number of information sources, volume of information, operational tempo and elevated levels of uncertainty, all which will place higher cognitive demands on operators. Thus it is critical that NCW research focus not only on technological innovations, but also the strengths and limitations of human-automation interaction in a complex system.

XX.1 INTRODUCTION

Network Centric Operations (NCO), also known as Network Centric Warfare (NCW), is a concept of operations envisioned to increase combat power by effectively linking or networking knowledgeable entities in a battlespace. Mission success is achieved by leveraging information superiority through a network, rather than through the traditional method of sheer numerical superiority through platforms and weapons. Key components of NCO include information sharing and collaboration which will promote shared situational awareness and overall mission success (DoD, 2001). To realize NCO, significant improvements will need to be made in areas of communications, sensor design, and intelligent automation. However, while technological advances are important for the successful integration of network centric operations, equally if not more critical is the need to understand how, when, where, and why the technology supports human decision makers and front line soldiers. Command and control domains are complex socio-technical domains in that technology is the means to an end (goal or mission) defined by human intentions. Advanced NCO technologies that are not designed with the express purpose of supporting military personnel in dynamic and uncertain situations with rapidly shifting goals are likely to fail.

The move from platform-centric warfare to NCW represents a shift in the role of humans both in mission planning and actual operation. As has already been evidenced in the development of fly-by-wire, highly automated aircraft and missile systems (such as Tomahawk and Patriot), military operators are less in direct manual control of systems, but more involved in the higher levels of planning and decision-making. This shift in control from lower level skill-based behaviors to higher level knowledge-based behaviors is known as human supervisory control (HSC). HSC is the process by which a human operator intermittently interacts with a computer, receiving feedback from and providing commands to a controlled process or task environment, which is connected to that computer (Figure 1).

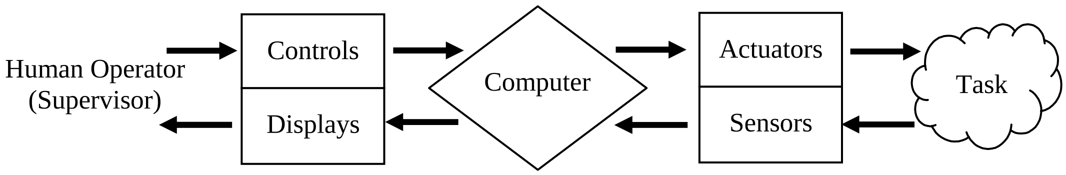

Figure XX.1: Human Supervisory Control (Sheridan, 1992)

HSC in military operations includes mission planning and passive and active intelligence operations, as well as missions involving manned aircraft, and unmanned air, ground, surface, and subsurface vehicles. The use of automated technologies and systems is a fundamental component of NCO; thus in the context of human interaction, NCO is a high level human supervisory control problem. The number and types of human-machine supervisory interfaces will expand accordingly in NCO, but there is little understanding of how functions should be allocated between humans and automation, what level of collaboration is needed, and how these can be supported with actual software and hardware. Given the influx of voluminous information sources in NCO, a particularly acute problem will be how to give operators enough information for well-informed decisions without reaching cognitive saturation. Moreover HSC problems in NCO are further complicated by the dynamic, uncertain, and time-pressured elements typical of command and control environments. Due to the increasing importance of HSC in NCO, the DoD has recognized that a lack of automation reliability and understanding of relevant HSC issues, as experienced both by individuals and teams, are among the primary barriers limiting exploitation of the full potential of NCO (DoD, 2001).

Using historical case studies as well as previous and current research studies, ten major human supervisory control issues have been identified as those HSC issues that are likely to cause degraded performance for both the system and the operators/decision-makers in futuristic network centric operations. They are:

- Information Overload
- Appropriate Levels of Automation
- Adaptive Automation
- Distributed Decision-Making and Team Coordination
- Mitigating Complexity
- Decision Biases
- Attention Allocation
- Supervisory Monitoring of Operators
- Trust and Reliability
- Accountability

These ten categories are not in rank order, are not mutually exclusive, and will often overlap theoretically as well as in actual design and testing. The importance of each category will be dependent on the mission context and relevant technological systems. They will now be discussed in turn in the following sections.

XX.2 TEN HUMAN SUPERVISORY CONTROL CHALLENGES IN NETWORK-CENTRIC OPERATIONS

XX.2.1 Information Overload

On March 28th, 1979, the worst US commercial nuclear power plant accident in history occurred at Three Mile Island in Middletown, Pennsylvania. The problem began when a main feedwater pump failure caused the reactor to automatically shut-down. In response to this, a relief valve opened to reduce the pressure of the system, but stuck open. There was no indication to plant *controllers that this had occurred. Due to the stuck valve, there was significant loss of reactor coolant water, subsequently causing the core of the reactor to overheat. There were no instruments that showed the level of coolant in the core, so it was thought to be acceptable based on the pressurizer coolant level. This lead to a series of human actions that made the problem worse, ending with a partial meltdown of the core. Operators were overwhelmed with alarms and warnings, numbering in the hundreds, over a very short period of time. They did not possess the cognitive capacity to adequately deal with the amount of information given to them during the unfolding events. Instead, they coped by focusing their efforts on several wrong hypotheses, ignoring some pieces of information that were inconsistent with their incorrect mental model.*

According to the DoD, the Global Information Grid (GIG), the actual information technology network that will link command and control agents, will be the enabling building block for NCO. The GIG is the end-to-end set of information capabilities, associated processes and personnel for collecting, processing, storing, and disseminating information to those who require it, on the battlefield or otherwise (DoD, 2001). Metcalf's Law states that the usefulness, or utility, of a network equals the square of the number of users (Shapiro & Varian, 1999). While this means forces will have access to exponential amounts of information over today's forces, it also means that information intake for the average NCO operator will be higher than ever seen been before in the command and control environment. Even if the information complexity does not increase (which is unlikely), mental workload will increase accordingly. The Yerkes-Dodson Law (Yerkes & Dodson, 1908), adapted to workload and performance (Figure XX.1) illustrates that beyond a task-dependent moderate level of arousal, individuals will become cognitively overloaded and their performance will drop. The problem is predicting when and how this overload will occur for a dynamic decision making environment, so that the amount of information any single person or group is required to process is manageable.

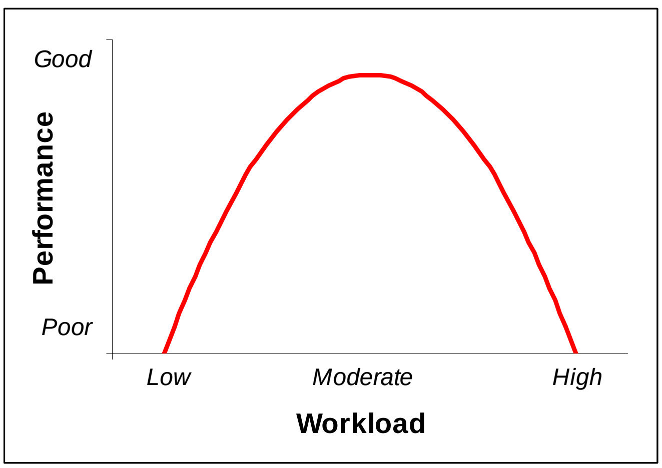

Figure XX.1: A Performance Adaptation of the Yerkes-Dodson Law

With the voluminous increase in incoming information in NCW, identifying the point or plateau of cognitive saturation is the key to designing systems that operators can effectively manage. Predicting this point of saturation is difficult for NCO systems because it is dependent on task, automation level, operational tempo, training, experience, and a number of other factors. Modeling and simulation can aid in predicting where these points are likely to be which could include cognitive, psychophysiological (to be discussed in a subsequent section), and predictive statistical models based on experimental simulations. An example of a predictive statistical modeling effort based on human-in-the-loop simulations that estimate the cognitive saturation point is the Tactical Tomahawk case study. Cummings and Guerlain (2004) demonstrated that in the futuristic control of multiple Tomahawk missiles, operators' performance and situation awareness (SA) was significantly degraded and overall task engagement time reached saturation when operators were required to control 16 missiles as opposed to 8 and 12 (Figure XX.2). These results were consistent across both low and high operational tempos and provide a visualization of Yerkes-Dodson in practice. Remarkably, air traffic control studies show similar results in that when asked to monitor and provide high level instructions to autonomously operated vehicles, beyond 16 aircraft (as with the missiles), performance significantly declines. While these results are somewhat narrowly scoped and not applicable to all areas of NCO, they do provide evidence that even when only intervening with high-level planning and supervision with no manual control, humans are limited in their capacity to effectively supervise multiple vehicles.

Development of an increased level of shared situation awareness and knowledge (SSAK) is major high level tenet of NCW (DoD, 2001), and is critical to realizing NCW's promise of substantial increases in combat power. However, the impact of information overload on situational awareness (SA), shared or otherwise, on human performance using an objective measurement of SA has been elusive. Situational awareness is generally thought to have three levels: 1) perception of important environmental cues, 2) comprehension of the situation, and 3) projection of future events and dynamics (Endsley, 1995). Direct system performance measures that use specific scenario manipulations to measure all three levels of SA are appropriate only in situations in which SA drives performance (Pritchett & Hansman, 2000). However, the correlation between SA and performance is debated (Pew, 2000). Direct experimental techniques, such as the SAGAT (Situation Awareness Global Assessment Technique) (Endsley, 1988), have been used in command and control research, but are not without problems. Questions remain as to whether SAGAT measurement techniques disrupt task performance, and if expectation of probes alters peoples' natural behaviors (Pew, 2000). Instead, subjective measures have traditionally been used in the human supervisory command and control domain, whereupon subjects self-rate their level of SA. It has been shown that these methods are of limited utility, providing more insight to judgment processes than situation awareness (Jones, 2000).

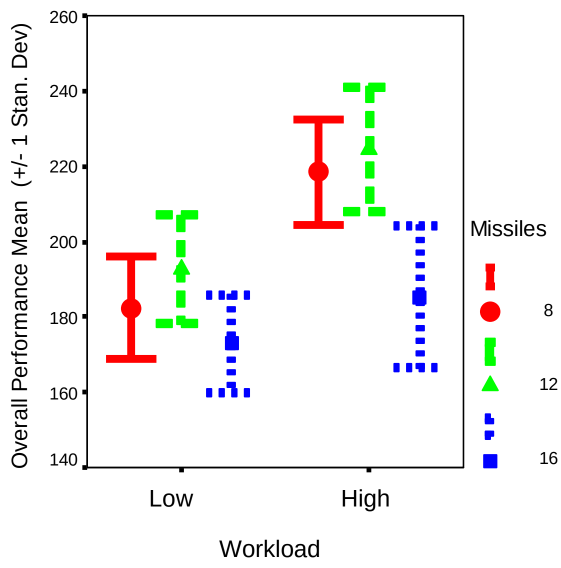

Figure XX.2: Supervisory Control of Multiple Missiles

XX.2.2. Appropriate Levels of Automation

The Patriot missile system has a history of friendly fire incidents that can at least be partially attributed to a lack of understanding of human limitations in supervisory control. On March 23rd, 2003, a RAF Tornado GR4 was shot down in friendly airspace by a Patriot missile. Two days *later, a USAF F-16 fighter pilot received a warning that he was being targeted by hostile radar, and fired a missile in self-defense. The "hostile radar" was in fact a Patriot missile battery aiming at him. On April 2nd, 2003, just nine days later, another Patriot missile shot down a US Navy F/A-18 returning to base from a mission. The Patriot missile has two modes: semi-automatic (management by consent – an operator must approve a launch) and automatic (management by exception – the operator is given a period of time to veto the computer's decision). However, in practice the Patriot is typically left in the automatic mode and the friendly fire incidents are believed to be a result of problems in the automatic mode.*

There are known "ghosting" problems with the Patriot radar in that because of operations in close proximity to other Patriot missile batteries, false targets will appear on a Patriot operator's screen. Under the automatic mode (management by exception), operators are given approximately 15 seconds to reject the computer's decision, which is insufficient to solve both false targeting problems as well as adequately address friend or foe concerns through any other means of communication. After the accident investigations, the US Army admitted that there is no standard for Patriot training, autonomous operations procedures (automatic mode) are not clear, and that operators commonly lose situational awareness of air tracks. Despite all the known technical and operational problems for Patriot, in the words of the US Army, "Soldiers [are] 100% reliant on Patriot weapon system (32nd Army Air and Missile Defense Command, 2003)."

While automating significant aspects of NCO is necessary so that information sharing can be both quick and comprehensive, what to automate and to what degree to automate a process/system is a central question in the design of NCO systems. Sheridan and Verplank (1978) outlined a scale from 1-10 where each level represented the machine performing progressively more tasks than the previous one, as shown in Table XX.1. Human interaction with automation represents a range of intermediate levels from 2-6 on this scale. For routine operations, higher levels of automation (LOAs) in general result in lower workload, while the opposite is true for low levels of automation (Kaber, Endsley, & Onal, 2000).

Table XX.1: Levels of Automation (Sheridan & Verplank, 1978)

Automation Level	Automation Description
1	The computer offers no assistance: human must take all decision and actions.
2	The computer offers a complete set of decision/action alternatives, or
3	narrows the selection down to a few, or
4	suggests one alternative, and
5	executes that suggestion if the human approves, or
6	allows the human a restricted time to veto before automatic execution, or
7	executes automatically, then necessarily informs humans, and
8	informs the human only if asked, or
9	informs the human only if it, the computer, decides to.
10	The computer decides everything and acts autonomously, ignoring the human.

It is possible to have a LOA too high or too low, each with its own distinct set of problems (Billings, 1997):

- HSC problems if LOA is too high
    - Manual or mental skill degradation.

    - Loss of situational awareness due to lack of automation transparency, complexity, and inadequate feedback.
    - More advanced automation issues such as brittleness & literalism; in other words, the automated system might not be able to handle novel or unexpected events. Moreover, it may not operate effectively in conditions near or at the edge of the intended operating envelope.
    - Time and difficulty to diagnose failures and manually take over.
- HSC problems if LOA is too low
    - Cognitive and working memory overload in routine tasks under time pressure.
    - Human decision biases and heuristics.
    - Lack of repeatability and consistency.
    - Complacency and boredom.
    - Greater human interdependency and chaos when something fails, unless safeguards are in place.

A significant design question in the context of developing HSC decision support systems for NCO is determining the appropriate LOA. Parasuraman et al. (2000) proposed that most tasks could be broken down into four separate information processing stages (information acquisition, information analysis, decision selection and action implementation), and that each could be assigned a level of automation separate and possibly different from the others. However, in the context of flexible human-automation interaction, subdividing a problem into these abstract stages may not go far enough. As proposed by Miller and Parasuraman (2003), each information processing task can be further divided into simple sub-tasks with differential levels of automation. For NCO, generalized cognitive tasks under these proposed information processing stages can be defined as:

- *Information acquisition*
    - Monitoring resources (such as friendly forces)
    - Monitoring systems (such as surveillance networks)
    - Communications
- *Information analysis*
    - Data fusion and display techniques
- *Decision selection*
    - Planning
    - Re-planning
    - Rapid resource allocation
- *Action implementation*
    - Individual vs. team interaction

Human supervisory control interactions with automation primarily fall under the analysis and decision processes (Figure XX.3). Of these two stages, information analysis is the one most affected by conversion of the military to network-centric principles, as the number and variety of available information sources to a robustly networked force is expected to increase dramatically. While the topic of information overload will be discussed in more detail in the next section, the potential for cognitive saturation in the NCO information analysis phase is likely to be caused by problems with data fusion. Data fusion in this sense is defined as the process by which raw information from disparate sources is filtered and integrated into relevant groupings before being displayed to users. Data fusion LOAs can vary such that low levels of automation for data fusion could include trend or predictive displays for single or multiple variables of interest, such as for tracking of enemy forces. Higher levels could include smart agents providing context dependent summaries of relevant information to users (Parasuraman et al., 2000).

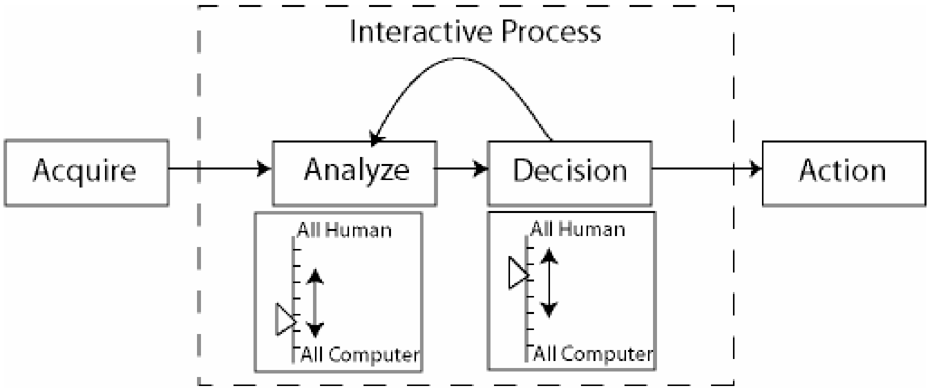

Figure XX.3: Human Supervisory Control Information Processing (Sheridan, 2002)

For example, Figure XX.4 represents two different levels of automation for the same data fusion task, which is the allocation of missiles to appropriate missions. On the left, operators are given detailed information about missile capabilities as well as mission requirements and constraints and the automation provides low level filtering (LOA 2, Table XX.1). On the right of Figure XX.3, operators are given the same information at a higher, more graphical level using a configural display, but a "smart" search algorithm narrows the solution space to nominally assist the operators (LOA 3, Table XX.1). While the graphical display may seem to be more intuitive, preliminary results with military personnel highlight a

Figure XX.4 Low vs. High Level Displays for Data Fusion significant problem with more advanced LOAs in general. While they generally agreed that low level display caused excessive workload, users felt that the high level graphical display did not provide enough detail sufficient to completely understand the problem (Cummings & Bruni, 2005). Designers of NCW systems will struggle with this same conundrum of selecting automation levels that both relieve operator workload yet provide enough information for operators to make informed decisions.

XX.2.3 Adaptive Automation

Research is underway to develop the Cognitive Cockpit which mitigates excessive pilot workload through augmented cognition (Diethe et al., 2004). The Cognitive Cockpit contains a Cognition Monitor that theoretically provides a real-time analysis of the cognitive–affective state of the pilot through unspecified psychophysiological measures as well as inferences about current attentional focus, ongoing cognition, and intentions. These inferences become inputs to a task manager that changes the levels of automation (as described in Table XX.1) for relevant tasking and workload. This process is termed Adaptive Dynamic Function Allocation, which increases automation levels in periods of high workload and vice versa. While adaptive automation is thought to benefit operators in periods of high workload, using a psychophysiological state as a trigger for adaptation is problematic due to the large amount of noise present in readings and extreme individual variability. While preliminary research in lab settings suggests psychophysiologic adaptive automation is possible, the current state of-the-art technology does not support operational use.

Military operations are often characterized by long periods of inactivity followed by intense periods of action during which time-critical decisions must be made. At these times performance is most critical, yet it will likely suffer due to the temporary information overload placed on the operator and the need for the operator to cognitively reorient to the new situation. With NCO and the emergence of a robustly networked force, the amount of information available to military personnel at all levels is exponentially greater. Therefore, the problem of information overload, particularly during brief bursts of actions, will become much more common.

One method to alleviate such problems is the use of adaptive automation (AA). Changes in the level of automation (Table XX.1) can be driven by specific events in the task environment, models of operator performance and task load, physiological methods, or by changes in operator performance (Parasuraman et al., 1992). AA has been shown to improve task performance (Hilburn et al., 1997; Prinzel et al., 2003), situational awareness (Kaber & Endsley, 2004) and lower workload (Prinzel et al., 2003). Two important questions to answer in deciding how and when to develop an adaptive automation decision support system are (1) when to use adaptive automation to determine under what circumstances the LOA should change, and (2) whether the computer or the human decides to change the LOA.

Specific cues in the environment used to change the LOA may be either time or event-related. For example, it may be determined that in order to maintain vigilance levels in a task, automation will be turned off for a certain period of time once an hour, forcing the operator to execute manual control during that period. Alternatively, it may be known that particular events in the environment will cause higher levels of workload than desired for a short time, during which the automation could increase to compensate. An example of this would be an operator responsible for multiple UAVs performing a bombing mission. It is well known that the cruise and loiter phases of flight are low workload, but that the approach and bombing phases require significant increase in operator workload. As a preemptive measure, adaptive automation could increase the LOA during the periods of high workload. This approach is problematic because it is scenario specific, will likely not handle unexpected situations correctly because desired changes in LOA must be pre-determined, and does not take into account operator variability. Some operators may have a much higher threshold for handling workload than others and thus may not require a change in LOA.

AA cueing can also be accomplished through models of operator performance which can predict the effectiveness of humans during particular processes and behaviors. Thus, the model's forecasted level of operator mental workload or performance on any number of specified tasks can be used to change the LOA. What is defined as an acceptable level of any measure predicted by the model must be carefully defined in advance. As defined by Laughery and Corker (1997), there are two main categories of human performance models: reductionist and first principles. Reductionist models break down expected behaviors into successively smaller series of tasks until a level of decomposition is reached that can provide reasonable estimates of human performance for these task elements. First principles models are based on structures that represent basic principles and processes of human performance. Performance models offer the advantage of flexibility in the sense that they can apply to a large range of situations, even unexpected ones, but often are costly and difficult to develop, especially if higher reliabilities are desired.

Psychophysiological measures such as the electroencephalogram (EEG), event-related brain potentials (ERPs), eye movements and electroculography (EOG), electrodermal activity (EDA), heart rate and heart rate variability (HRV), breathing rate and blood pressure have all been correlated with mental workload to varying degrees of success. Experimentally, these methods are advantageous because they are not task specific and they can continuously record data. One significant problem is that often the devices used to take these measurements are obtrusive and physically uncomfortable for subjects, creating a possible anxiety effect. While this is still a barrier to overcome, technologies such as the LifeShirt®, wireless EEG sensor headsets (Berka et al., 2004) and stereo cameras for tracking eye movements without the use of headgear can provide more ecological measurements of psychophysiological metrics. Other significant problems with psychophysiological measures include the large amount of noise present in readings, and extreme individual variability. One way to lessen these effects is to use combinations of measurements taken in concert, as done by Wilson and Russell (2003). They showed accuracies in excess of 85% classifying operator states in real-time using artificial neural networks trained on a battery of 43 physiological features. However, while psychophysiological measures have been used to adaptively allocate automation functions in research environments, because of the previously discussed limitations, transferring experimental AA to operational AA has not yet been demonstrated.

Of all possible psychophysiological measures, analysis of brain waves has historically been the primary method of investigating neural indexes of cognition (Fabiani, Gratton, & Coles, 2000). Numerous studies have successfully used EEG measures of workload to discriminate between differences in operator attention and arousal (Berka et al., 2004; Kramer, 1991; Pope, Bogart, & Bartolome, 1995). They used engagement indexes based on the ratios of different EEG bands (alpha, beta, theta, etc), alpha suppression, or increased beta levels to detect changes in workload. However, as Prinzel et al. (2003) notes, EEG-based systems are only able to measure gross overall changes in arousal, not different types or finer levels of cognitive load. In contrast, the P300 component of ERPs has been associated with the availability of information processing resources, varying in amplitude as a function of primary task load, and its latency affected by stimulus evaluation time (Polich, 1991; Rugg & Coles, 1995). The P300 has been documented as an effective measure of mental workload (Donchin, Kramer, & Wickens, 1986; Kramer, 1991). Despite this, use of ERPs in non-laboratory settings to measure workload in real-time has proven difficult, as they are obtained from averaging of EEG signals over a number of trials. Methods for addressing this shortcoming are under development, but Humphrey and Kramer (1994) were able to discriminate between different workload levels 90% of the time using only 1 to 11 seconds of ERP data (approximately 1-11 trials).

Eye movements have also been used extensively in a variety of studies. They have been used in distraction studies to identify areas of attention, but measures of blink duration, frequency and pupil diameter have been correlated with visual and cognitive workloads. In general, eye blink duration and frequency decrease as both visual and/or cognitive workload increases (Hankins & Wilson, 1998; Orden et al., 2001; Veltman & Gaillard, 1998). Though pupil diameter is affected by ambient light levels, stimulus perception, and habituation, the degree of pupillary dilation has been shown to increase with higher cognitive loads (Andreassi, 2000; Hess & Polt, 1964; Orden et al., 2001). Specific characteristics of eye movements, such as fixation duration, dwell times and saccade durations are task-dependent and thus are extremely difficult to tie to changes in mental workload in a generalizable way. However, Simon et al. (1993) demonstrated that the more general visual behavior of scanning patterns becomes more organized when task difficulty increases.

Finally, AA may be based upon performance-based measures, whereupon some performance metric such as reaction time or task accuracy is used to determine mental workload. While generally easier to measure and quantify than physiological measures, performance measures are generally task-specific (and thus not generalizable to other tasks) and often require the subjects to modify their natural task behavior to accommodate the experimental objectives. Moreover, performance-based measures also may only give the experimenter discrete samplings of operator workload at specific intervals instead of a constant measurement. This could be inappropriate for some applications characterized by rapid changes in the environment like what will likely occur in high operational tempo settings in NCW, necessitating quick switches between automation modes.

XX.2.4. DISTRIBUTED DECISION-MAKING AND TEAM COORDINATION

In 1994, Operation Provide Comfort provided humanitarian aid to over one million Kurdish refugees in northern Iraq in the wake of the first Gulf War. As part of this, the US sought to stop Iraqi attacks on the Kurds by enforcing a no-fly zone. The no-fly zone was patrolled by USAF fighters (F-15s), supported by airborne warning and control (AWAC) aircraft. On April 14, 1994, two US Army Black Hawk helicopters were transporting U.S., French, British, and Turkish commanders, as well as Kurdish para-military personnel across this zone when two US F-15 fighters shot them down, killing all 26 on board. The Black Hawks had previously contacted and received permission from the AWACs to enter the no-fly zone. Yet despite this, AWACs confirmed that there should be no flights in the area when the F-15s misidentified the US helicopters as Iraqi Hind helicopters. The teamwork displayed in this situation was a significant contributing factor to the friendly fire incident, as the F-15s never learned from AWACs that a friendly mission was supposed to be in the area. It was later determined that the F-15 wingman backed up the other F-15's decision that the targets were Iraqi forces despite being unsure, which was yet another breakdown in communication. Each team member did not share information effectively, resulting in the distributed decision making of the AWACs and F-15s pilots to come to incorrect and fatal conclusions.

Platform-centric command and control (C^2) in past military operations often avoided distributed decision-making and minimized team coordination in favor of a clear hierarchy for both information flow and decision making. Many decisions were made by a select few at the top level of command, and pains were taken to decompose various factions of the military into small, specialized niches that had little direct contact between one another (a hierarchical waterfall approach). This has begun to change in recent times, and a fully realized vision of NCW will require that both local and global teams of distributed decision makers, not a few people at the top of a hierarchy, make decisions under time-pressure. Therefore, understanding the issues unique to team-based coordination and decision-making take on new importance in the context of NCW. The question is how to make effective decisions between and within distributed teams, particularly in the complex, data-rich, and time-compressed situations often seen in military C^2 and NCW scenarios.

A fundamental building block of good decision making is a high level of situation awareness (SA) (Endsley, 1995), and in the case of distributed decision-making a high level of team SA or shared situation awareness (SSA). As previously discussed, the three levels of individual SA are: 1) perception of important environmental cues, 2) comprehension of the situation, and 3) projection of future events and dynamics (Endsley, 1995). Team SA involves individual SA for each team member, plus the SA required for overlapping tasks and team interactions (Endsley, 1995). Endsley and Jones (2001) expanded upon this definition by outlining a layered model of how teams achieve high levels of team SA. They detail what constitutes SA requirements in a team setting at each previously defined level (Table XX.2), the devices and mechanisms used to achieve shared SA, and SA processes that effective teams use. Team SA devices include spoken and non-verbal communications, visual and audio shared displays, and a shared environment. Given the need to support team SA in a dynamic decision-making environment under significant uncertainty typical of NCW settings, more research is needed in three areas: 1) What technologies can support NCW team decision making and performance, 2) What NCW team processes, both locally and distributed, are effective, and 3) What is the impact of different team architectures for NCW operations?

Table XX.2. Team SA Requirements for Shared Information (Endsley & Jones, 2001)

Level 1 SA - Data
System
Environment
Other team members
Level 2 SA - Comprehension
Status relevant to own goals/requirements
Status relevant to other's goals/requirements
Effect of own actions/changes on others
Effect of other's actions on self and overall goal
Level 3 SA - Projection
Actions of team members

There has been extensive work in the computer-supported cooperative work (CSCW) community that examines how different types of technologies, to include both software and hardware, support effective team decision making. Of particular interest to NCW is the relationship between collaborative technologies over space and time as depicted in Figure XX.6 (adapted from (Johansen, 1988)). Developing technologies that promote collaboration both locally and remotely in space as well as synchronous and asynchronous in time is highly relevant to NCW. For example, Navy personnel are assigned to different watch sections for a ship's power plant, thus share a space but communicate across different times. One collaborative technology they use to pass knowledge is a log book. Logs are used by watchstanders to record significant events, time they occurred, who was notified, what actions were taken, etc. While they have existed on paper for hundreds of years on ships, written logs are now giving way to electronic logs, with added benefit of an easier search space and automated reminders.

However, while many technologies developed for corporate settings show promise for NCW applications (e.g., electronic whiteboards (Price et al., 2001), table top displays (Scott, Carpendale, & Habelski, 2005)), more research is needed into both the promised benefits, as well as unintended consequences. For example, chat, a popular and ubiquitous synchronous communication tool for remotely distributed military individuals and teams, can have unintended consequences in terms of degraded human and system performance (Cummings, 2004c and see discussion XX.2.7.) In another study focusing on the benefits of shared displays between co-located team members, the shared displays unexpectedly contributed to degraded performance due to an increase in workload (Bolstad & Endsley, 2000). Given the high risk and uncertainty of military time sensitive operations, more investigation into the impact of new collaborative technologies is warranted.

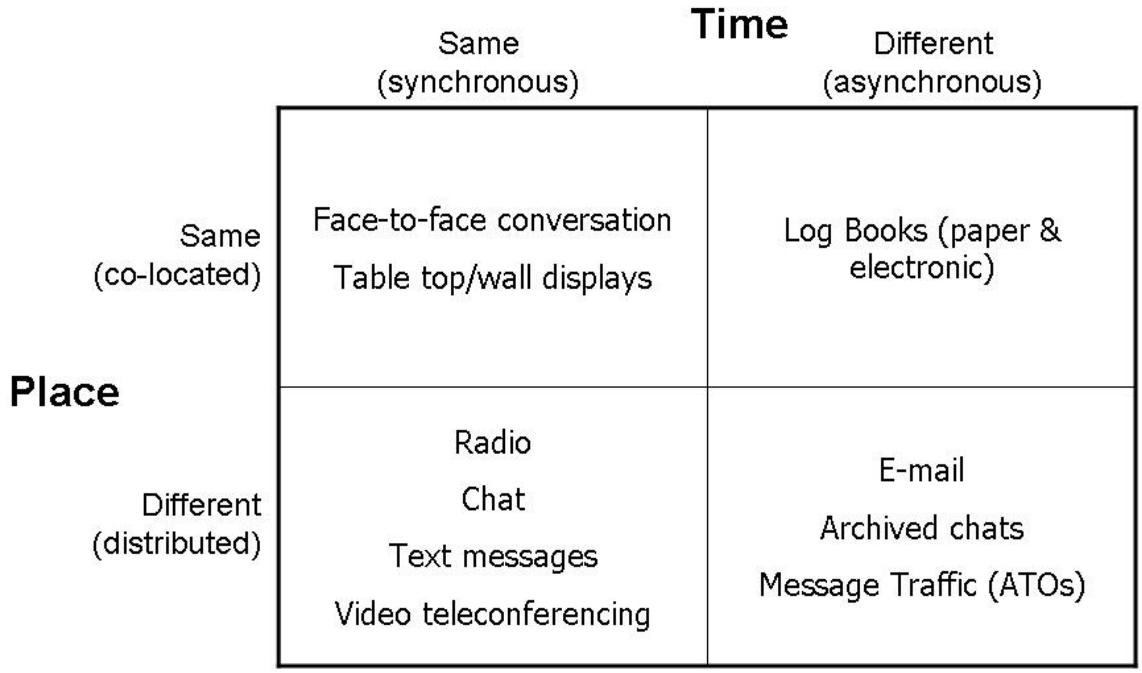

Figure XX.6: NCW Time and Place Collaborative Matrix

Understanding team processes, the second area of import for distributed decision making in NCW settings, is yet another area that can benefit from research already underway in the CSCW community. However, the caveat remains that military settings carry unique constraints not experienced in business settings. Processing capabilities of humans are limited, especially in time critical domains, so distribution of information, resources and tasks among decision-makers must be done in such a way that the cognitive load of each person on a team is not exceeded. However, as pointed out by Hutchins in aviation and ship navigation (Hutchins, 1995a, 1995b), cognition is distributed across persons and artifacts in a system such that teams often possess shared capabilities beyond the sum of individual contributions..

While difficult to capture, Cooke et al., (2003) measured distributed cognition in a command and control setting through team knowledge and succeeded in predicting subsequent team performance. However, distributed cognition across large scale time sensitive operations with multiple entities, both human and automated, is not well understood as well as how the introduction of new technologies supports or detracts from team SA in NCW. Boiney (2005) echoes this sentiment and in terms of time sensitive targeting, highlights the need for better understanding of collaborative sensemaking, the establishment of trust, distributed team SA, and appropriate information sharing to include communications and awareness cueing.

Lastly, design of team architectures to include role allocation, team geographic distributions, communication, and team sizes is critical to successful NCW distributed decision making and team coordination. Research has shown that organizations operate most efficiently when their structures and processes match their mission environments (Levchuk et al., 2002), but it is not always clear whether traditional top-down hierarchies or more lateral heterarchies provide the most efficient structure (Baker et al., 2004). Moreover, sensor quality and operational tempo can drive the need for distributed versus centralized control (Dekker, 2002), thus further complicating the team architecture problem by adding a technological artifacts.

The problem of uncertainty, driven both by the lack of sensor as well as human information, will likely be a significant driver of NCW team decision-making success or failure. Price et al., (2001) hypothesize that teams organized by functional structure (a team than specializes in performing a task) perform better in terms of time and task accuracy when there is a high degree of certainty in the environment. When an environment is characterized by uncertainty and unpredictability, Price et al, (2001) assert that instead of a functional structure, teams should be organized through a divisional structure that is relatively autonomous and independent. Unfortunately the very nature of NCW is high unpredictability and uncertainty, but since information sharing between teams is critical, they cannot be either independent or autonomous, thus the characterization of teams either as functional or divisional in NCW is problematic.

Alberts and Hayes (2003) offer one solution for the optimal team architecture problem with their proposed 'edge organizations', which are characterized by widespread information sharing and peer-to-peer relationships, but their design offers very few details on how to actually implement them. Levchuk et al., (2004) propose that the most effective team architecture will be a hybrid organization which utilizes the beneficial characteristics of hierarchies as well as heterarchies such that control is an emergent property which is a function of the environment, scenario initial conditions, and adversary behavior.

XX.2.5. COMPLEXITY MEASURES

On December 20th, 1995, American Airlines flight 965 was enroute to Cali, Colombia from Miami, Florida. Extremely behind schedule, as it approached Cali, the flight crew accepted a modified, unplanned and unfamiliar arrival route. Trusting the flight management system (FMS) to automatically come up with the next correct waypoint after having only entered the first letter of it, the captain mistakenly chose the computer's first choice, an incorrect waypoint 132 miles northeast of Cali. This resulted in a wide turn to the east, whereas the flight crew knew that they were supposed to be on a straight approach. Recognizing something was wrong, but not exactly what, the flight crew took manual control, turning right again towards Cali. Unfortunately, the approach into Cali was surrounded by mountains and flight 965 was now sufficiently off course that they flew into the side of a mountain, killing all but 4 aboard the Boeing 757. A clear contributor to this accident was the automation bias displayed by both the captain and first officer. They continued to rely on FMS-assisted navigation even when it became confusing and cognitively demanding during a critical segment of flight.

Information complexity is a growing problem in many domains, with particular applicability to NCW. Complexity will be impacted by both the amount and sources of information, and will be further exacerbated in the future as sensor technologies improve and the volume of available data continues to grow. NCW operators will be required to understand resultant critical relationships and behaviors of that data at the same or higher level than today. Increased complexity will usually manifest in increased workload and/or unpredictability of the system, to include the human, which will have a negative effect on human and system performance (Miller, 2000). There are also well known limitations to human performance (see XX.2.1) that are likely to be encountered as workload increases. Therefore, it is important that the interfaces NCW operators interact with help to reduce and manage this increased level of data complexity.

Complexity, as defined by Merriam-Webster, is "the quality or state of being hard to separate, analyze, or solve". The use of the term 'hard' implies that complexity is relative, which was captured by Miller (2000) when he described the difference between actual and perceived complexity. Perceived complexity results from those elements of a task or situation that make it hard to deal with, or in other words, what makes a system seem difficult. Actual complexity implies the use of more objective criteria for complexity, and does not take into account humans' perceptions of a task or situation.

Perceived complexity can be divided into three general dimensions: 1) Component complexity: the number and diversity of components, 2) Relational complexity: the number and diversity of links between components, and 3) Behavioral complexity: the number and diversity of behaviors system components can exhibit (Miller, 2000). Often, these dimensions are not independent and changes in complexity are driven by interactions between them. For example, a common operator task in current military operations is tracking and interacting with multiple entities such as aircraft, and in the future, multiple heterogeneous unmanned vehicles. An increase in the number of entities to monitor increases component complexity, but it is likely that the proximity of these new entities to the existing ones will also be important (an increase in relational complexity). It is also possible that new tracks will exhibit different behaviors than existing tracks (an increase in behavioral complexity), which will be particularly problematic in the future vision of “swarming” autonomous vehicles (Cummings, 2004b).

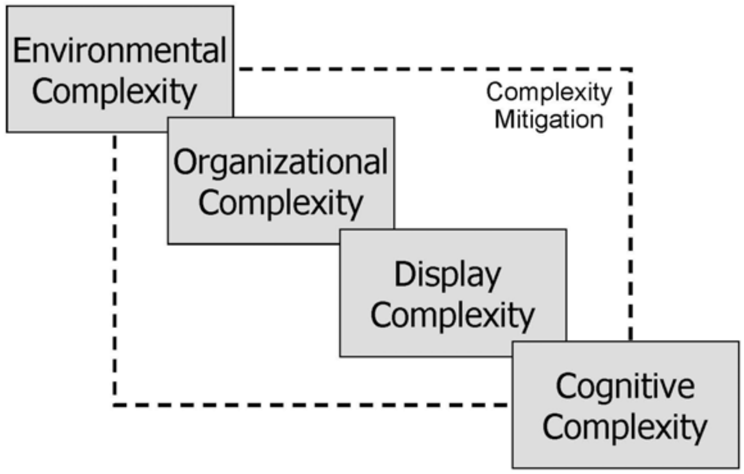

Figure XX.7: The Complexity Chain for Human Supervisory Control

The gap between actual and perceived complexity is represented in Figure XX.7. For human supervisory control tasks, there is an objective, actual complexity of the world generated by the environment (i.e., the number of vehicles to control, operational tempo, weather conditions), which is separate from the perceived, or cognitive, complexity of operators. Operators’ perceived complexity can be mitigated through organizational and/or displays strategies. Organizations like the FAA and DoD will institute policies and procedures to mitigate environmental complexity such that operations are both safe and meet goals. For example, separation standards exist to mitigate complexity for air traffic controllers which promoted safety. However, when airspace becomes saturated, the need to maintain these organization-driven constraints causes situation complexity, and thus workload, to increase.

In human supervisory control domains, displays are nominally designed to further reduce complexity by representing the environment so that a correct mental model can be formed and correct interactions can take place (Woods, 1991). Thus displays should reduce complexity and hence workload through transforming high-workload cognitive tasks such as mental computations into lower workload tasks through direct perception, i.e. visually (Miller, 2000). In addition, displays should promote effective cognitive strategies such as grouping and rule formation to further reduce perceived complexity. However, one drawback to new display technology is that in complex and dynamic human supervisory control domains such as ATC and NCW, it is not always clear whether a decision support interface actually alleviates or contributes to the problem of complexity. As illustrated in the Cali tragedy, the FMS, which was meant to reduce environmental complexity and pilot workload, actually was a significant contributor to the accident. While preliminary research indicates that environmental, organizational, and display complexity from Figure XX.7 can be measured as separate constructs and that display elements can add to cognitive complexity (Cummings & Tsonis, 2005; Cummings, Tsonis, & Cunha, 2005), a more principled approach is needed in the development of metrics for display and organizational complexity.

One other potential problematic area for cognitive complexity that has emerged from recent display technological advances is the aesthetic seductiveness of compelling visual, aural, and haptic displays, otherwise termed as the “cool factor.” While displays that use advanced technologies such as 3D, holographic images, virtual reality, and layering invariably elicit a “that’s cool” response, emerging trends in research show that not only are these technologies often not helpful, they can also be harmful. For example, 3D displays are gaining in popularity, yet in recent command and control studies, realistic 3-D perspective views produced poor performance for precise relative position and distance judgments.

Furthermore, despite the fact that operators preferred 3-D icons, conventional 2D symbols produced superior performance (St. John et al., 2001). Smallman et al., (2005) attribute this disparity between operators' preference for 3D displays and their sub par performance to a concept they term "Naïve Realism." Operators naïvely prefer displays that mimic realistic scenes over representations that are flawed and imprecise, which could lead to poor performance.

Multi-modal displays have become increasingly popular in military NCW settings (e.g., (Osga et al., 2002; Shilling et al., 2003) These types of displays will often incorporate multiple sensory inputs and outputs in addition to visual information including speech-to-text, spatial audio, touch displays, haptic feedback. While humans can successfully integrate information from multiple channels, it is not clear that such "cool" technologies actually help operators. In addition, virtual reality simulating face-to-face communications for remotely located military teams has not been shown to be effective (Wells & Hoffman, 1998) and display layering has not yet produced any measurable benefit (Bolia, Nelson, & Vidulich, 2004). While these types of displays and "cool" software applications like animation (see Tversky, et. al, (2002) for a review) superficially seem as if they aid operators in complex decision-making tasks, significantly more attention is needed to the role these technologies play in adding to cognitive complexity as well as the overall cost-benefit analysis for network centric operations.

XX.2.6. DECISION BIASES

On July 3rd, 1998, 290 passengers and crew took off from Bandar Abbas airport in Iran on Iran Air Flight 655, bound for Dubai in the United Arab Emirates. Tragically, it never made it as the Aegis class cruiser USS Vincennes shot down the flight over the Strait of Hormuz, killing all on board. While many factors contributed to this accident, such as the tense atmosphere in the Gulf at that time due to the Iran-Iraq war, the root cause can be attributed to the complexity associated with USS Vincennes' advanced tracking radar. It was designed in the 1980s for open water battles with the Soviet Navy, and as such, was capable of tracking hundreds of missiles and airplanes simultaneously. Efforts that might have mitigated the high level of complexity of this system were not undertaken in favor of improving the detection capabilities of the system. Two pieces of wrongly interpreted data resulting from the complexity of the human-machine interface caused the ship's commander to make the erroneous decision to fire on flight 655. First, flight 655 was reported as decreasing in altitude when it was in fact doing the opposite, representing that it was not in an attack profile. Second, the flight's Identification Friend or Foe (IFF) signal, designed to differentiate between civilian and military aircraft, was misidentified as being Mode II (military) instead of Mode III (civilian).

A defining characteristic of NCW is the expected increased information-sharing tempo over platform-centric forces of the past, which will require rapid decision-making with imperfect information. Humans in general, and especially under time pressure, do not make decisions according to rational decision theories. Rather they act in a naturalistic decision-making (NDM) setting in which experience, intuition, and heuristics play a dominant role (Klein, 1989). Humans generally employ heuristics in order to reduce cognitive load (Tversky & Kahneman, 1974; Wickens & Hollands, 2000), which will likely be the case in NCW settings. However, while heuristics can be useful, powerful tools, they can also introduce bias in decision making, especially when coupled with large amounts of information time pressures. The Vincennes accident illustrates how the three general classes of heuristics, representative, anchoring, and availability, could be problematic in NCW.

In the representative heuristic, probabilities are evaluated by the degree to which A resembles B. This heuristic can provide the illusion of validity because decision makers are generally insensitive to prior probabilities. In the case of the Vincennes, the operators ignored the likely prior probability that the plane was a commercial airliner which traveled on a fairly regular schedule and instead believed the much less likely event that the plane was a military fighter. Misconception of chance is a classic sign of the representative heuristic (Tversky et al., 1974).

The anchoring heuristic, an initial guess or hypothesis that is not adjusted appropriately given new information, was also evident in the Vincennes incident. The initial hypothesis was made that the radar contact was an enemy fighter and despite evidence that the plane was climbing and not descending (which would not be the case if the aircraft was attacking) in addition to published commercial airline schedules, all additional incoming information was discounted. Time pressure exacerbated this anchoring of the initial hypothesis, which will likely be a significant problem in NCW.

Lastly, the availability heuristic in which probabilities of events are judged based on recency and retrievability was also evident in the Vincennes accident. One year prior, the USS Stark had been hit by two Exocet missiles in the Persian Gulf, fired from a relatively low, fast flying Iraqi Mirage. In addition, on the morning of the shoot down, the Vincennes had engaged with Iranian gunboats so tensions both locally and globally were high. Given human memory limitations and the known propensity for humans to resort to often flawed applications of heuristics, a significant challenge in the design of NCW decision support systems will be how to give operators timely, probabilistic, and unbiased information.

While humans can be effective in naturalistic decision making scenarios in which they leverage experience to solve real world ill-structured problems under stress (Zsambok, Beach, & Klein, 1992), as previously discussed, they are prone to fallible heuristics and various decision biases that are heavily influenced by experience, framing of cues, and presentation of information. For example, confirmation bias takes place when people seek out information to confirm a prior belief and discount information that does not support this belief (Lord, Ross, & Lepper, 1979). Another decision bias, assimilation bias, occurs when a person who is presented with new information that contradicts a preexisting mental model, assimilates the new information to fit into that mental model (Carroll & Rosson, 1987). Of particular concern in the design of intelligent decision support systems that will support NCW processes is the human tendency toward automation bias, which occurs when a human decision maker disregards or does not search for contradictory information in light of a computer-generated solution which is accepted as correct (Mosier & Skitka, 1996; Parasuraman & Riley, 1997). Operators are likely to turn over decision processes to automation as much as possible due to a cognitive conservation phenomenon (Fiske & Taylor, 1991), and teams of people, as well as individuals, are susceptible to automation bias (Skitka & Mosier, 2000). Human errors that result from automation bias can be further decomposed into errors of commission and omission. Automation bias errors of omission occur when humans fail to notice problems because the automation does not alert them, while errors of commission occur when humans erroneously follow automated directives or recommendations (Mosier et al., 1996).

Many studies have demonstrated clear evidence of automation bias in laboratory settings. Layton, Smith, and McCoy (1994) examined commercial pilot interaction with automation in an enroute flight planning tool, and found that pilots, when given a computer-generated plan, exhibited significant automation over-reliance causing them to accept flight plans that were significantly sub-optimal. Skitka, Mosier, and Burdick (1999) found that when automated monitoring aids operated reliably, they led to improved human performance and fewer errors as opposed to not having an aid. However, when the automation failed to detect and notify operators of an event, or incorrectly recommended action despite the availability of reliable confirmatory evidence, human error rates increased significantly. More directly related to NCW processes, Cummings (2004) found evidence of automation bias when an intelligent decision aid recommended a single course of action for retargeting missiles to emergent targets.

Time pressure was a component of all of these laboratory studies, and there is unfortunately ample anecdotal evidence in the “real world” of automation bias under time pressure where the consequences were deadly. The Cali accident mentioned in section XX.2.5 was caused in part due to overtrust and automation bias in the FMS recommendations. In terms of NCW settings, the Patriot case study discussed in section XX.2.1 was also due to automation bias. Given the laboratory evidence that given an unreliable system, humans are still likely to approve computer-generated recommendations, it is not surprising that under the added stress of combat, Patriot operators did not veto the computer’s solution.

As these cases demonstrate, heuristics and decision biases are a very real concern in the development of intelligent systems that provide decision support for humans, especially those in time critical domains. Designers of intelligent systems should be aware of the potential negative effects on decision making in terms of inappropriate heuristics and biases as the human is further removed from the decision control loop. Intelligent decision aids are intended to reduce human error and workload but designers must be mindful that higher levels of automation combined with unreliable systems can actually cause new errors in system operation if not designed with human cognitive limitations and biases in mind. Design of an intelligent system that provides decision support must consider the human not just as a peripheral device, but also as an integrated system component that in the end, will ultimately determine the success or the failure of the system itself.

XX.2.7. ATTENTION ALLOCATION

Instant messaging was a primary means of communication between Navy ships during Operation Iraqi Freedom in 2003. While instant messaging, otherwise known as chat, has many advantages for rapid response in critical time-pressure command and control situations, operational commanders have found it difficult to handle large amounts of information generated through chat, and then synthesize relevant knowledge from this information (Caterinicchia, 2003). Chat enables faster responses in time critical command and control situations, but as described by the military, managing and synthesizing the large amount of information transmitted can sometimes be a "nightmare". The addition of instant messaging in the command and control loop requires a division of attention from the primary task, which may not always be appropriate. If the power of an intelligent automated and adaptive agent was harnessed so that the computer could determine more optimal scheduling patterns for the presentation of instant messages, it is possible that information would not be a detrimental interruption and overall human performance improved (Cummings, 2004c).

An important task in supervisory control is often one of how to allocate attention between a set of dynamic tasks. In deciding on an optimal allocation strategy, the operator acts to balance time constraints with relative importance of the required tasks. Due to the expected increases in the number of available information sources in NCW, volume of information and operational tempo, greater attentional demands will be placed on operators. This is a fundamental and critical HSC problem in NCW. There are two general areas where attention allocation issues are likely to occur in NCW: Preview times/stopping rule generation and primary task interruption.

XX.2.7.2.1. Preview Times and Stopping Rules

In NCW, the problem of attention allocation issues and preview times occurs when an operator expects sensor information at established time intervals to accomplish some task, and must act on this information , whether complete or not, before a deadline. For example, an air defense warfare coordinator (AWC) on a Navy ship could be responsible for several tasks: Identifying unknown air tracks as friendly, enemy or commercial air, monitoring these identified tracks, providing warnings to enemy aircraft within a certain radius, and providing launch orders for additional defensive aircraft against encroaching enemies. Each of these tasks could involve numerous sub-tasks such as air traffic communications, visual confirmations, etc. Obviously some tasks are more important than others; shooting down threatening enemy aircraft is higher priority than tracking a commercial air flight. Enemy and commercial air flight launches or a sudden re-classification of a track could represent unpredictable increases to task load. Additionally, the AWC receives information updates only at discrete intervals as the radar sweeps by an area of interest, or scheduled transmissions from equipment are received. Thus the AWC operator expects information to arrive in a certain time interval that could reduce uncertainty, but is sometimes faced with time-critical decisions that may or may not be able to wait for this information.

A central issue with the concept of preview times is how to maintain task priority when additional information is expected in the future, and how emergent situations influence an operator's ability to assimilate this preview information. Tulga and Sheridan (1980) investigated some aspects of this in a generic multi-task supervisory control paradigm. They found that at high workloads, the time subjects planned ahead was inversely proportional to the inter-arrival rate of new tasks. Using a similar paradigm, Moray et al. (1991) found that even if subjects were given an optimal scheduling rule, they were unable to implement it under enough time pressure, resorting instead to significantly non-optimal heuristic rules. In both experiments however, it was not possible to gain new information about specific tasks that would influence planning, nor were there unexpected events that significantly changed the nature of the task. In recent work examining intelligent agent predictions for future periods of high workload in order to aid operators controlling multiple UAVs, results revealed that subjects fixated on attempts to globally optimize an uncertain future schedule to the detriment of solving certain, local problems (Mitchell, Cummings, & Sheridan, 2005). It is clear from these initial efforts that more research is required to understand the effects of preview times, especially with information updates and unanticipated occurrences.

A related issue to preview times is that of stopping rule generation. Stopping rules are the criteria that individuals use to “satisfice” in uncertain situations, i.e. choosing the current best plan that is good enough. The general problem is as follows: Say that an operator has initial information, such as locations of friendly and enemy forces, and incoming information of various reliabilities and different times of arrivals, such as updates on enemy movements from voice communications and satellite images. The longer operators wait to make a decision on what to do with their forces, the more information they can gather (though not necessarily better due to its probabilistic nature), but they have a time limit in which to act. An individual’s stopping rule would determine when the decision was made. Another interesting issue would be to observe how and if initial decisions were changed as more information was received and final time deadlines approached. A better understanding of the relationship between stopping rules and preview times in NCW is needed because by its very nature, NCW hinges upon successful information sharing. However, due to the stream of data from multiple sources and the need for rapid decisions, operators will have to weigh the benefits of gathering more information that will reduce uncertainty against the cost of a possibly delayed decision.

XX.2.7.2.2. Primary Task Interruption

Instead of a situation where an NCW operator has multiple dynamic tasks that vary in priority, consider a common scenario where the operator has a well-defined primary task along with secondary tasks such as flying a UAV while also managing the weapons systems. In time-pressure scenarios, interruptions of a primary task caused by secondary tasks can increase mental processing time and induce errors in the primary task (Cellier & Eyrolle, 1992). For supervisory control tasks such as command and control or monitoring of displays, operators spend time monitoring unfolding events which may or may not be changing rapidly, and they also will periodically engage in interactive control tasks such as changing the course of UAVs or launching a missile. When task engagement occurs, operators must both concentrate attention on the primary task, but also be prepared for alerts for external events. This need to concentrate on a task, yet maintain a level of attention for alerts, causes operators to have a conflict in mental information processing. Concentration on a task requires “task-driven processing” which is likely to cause decreased sensitivity or attention to external events. Interrupt-driven processing, needed for monitoring alerts, occurs when people are sensitized and expecting distraction.

While interrupt and task driven processing can be present in a person, attention must be shared between the two and switching can incur cognitive costs that can potentially result in errors (Miyata & Norman, 1986). The conflict between focusing on tasks and switching attention to interruptions is a fundamental problem for operators attempting to supervise a complex system which requires dedicated attention but also requires operators to respond to secondary tasks, such as communications or alerts from non-critical sub-systems. In addition, Gopher et al. (1996) demonstrated that not only is there a measurable cost in response time and decision accuracy when switching attention between tasks, but costs are also incurred by the mere reconsideration of switching tasks. Moreover, often what appears to be an innocuous peripheral, secondary display feature such as scrolling of text in a chat window can have negative consequences because the distraction requires cognitive effort in considering whether or not it needs attention (Maglio & Campbell, 2000; Somervell et al., 2001).

McFarlane (1999) proposed that of the four ways of coordinating user-interruption which are 1) immediate, 2) negotiated, 3) mediated, and 4) scheduled, people perform better when they negotiate the onset of interruptions. However, one consequence is that they may not be prepared to effectively handle unplanned interruptions. Additionally, when forced to handle interruptions immediately, operators will execute relevant tasks promptly but performance can decline with an increase in mistakes. Altmann and Trafton (2002) have proposed that the interruption lag, the time between the interruption alert and the interruption event, is an intervention window that can be used reduce task resumption time. However, the use of cues during the interruption lag has not been shown to decrease recovery time, and in one study, appeared to increase recovery time (Altmann & Trafton, 2004; Miller, 2002).

While these studies addressed mitigation of detrimental interruption effects through process, few have investigated uses of technology to this end. McFarlane (2002) proposed an intelligent aiding system to both prioritize incoming interruptions as well as aid operators in task recovery through voice communication with the computer such as “What was I last working on?” St John et al., (2005) designed a display table that logs changes during the interruption to help users recover from interruptions in a dynamic monitoring task, which appeared to be superior to raw video playback of an interrupted task. These studies have only uncovered the tip of the iceberg for interruption mitigation design strategies. Because of the high likelihood and potential cost of induced errors due to interruptions and switching costs in NCW supervisory control systems, much work remains to be done in this area.

XX.2.8. SUPERVISORY MONITORING OF OPERATORS

October 29, 1998, two Boeing 737s were on standard air routes from Darwin to Adelaide and Ayers Rock to Sydney, respectively, at the same flight level of 37,000 feet. They were scheduled to conflict with each other, so protocol dictated that a 2,000 feet vertical separation standard be applied. This was noted by both the air traffic controller and a supervisor assigned to that particular sector 90 minutes before the conflict actually occurred, and marked for later action. In the next 90 minutes, traffic levels steadily increased and a third air traffic controller began to assist the other two already working in the conflict sector. The third controller assumed the coordinator position and attempted to deal with any items that seemingly required attention as he attempted to gain some idea of the traffic disposition. Despite the addition of a third coordinating ATC controller and the previous identification of the conflict, the pending conflict was subsequently overlooked. Instead, a violation of the minimum vertical separation distance occurred as one of the aircraft in question alerted ATC of the conflict at the last minute. This was an instance where the supervisor failed to detect a major failure of his supervisee, despite indications that one might occur.

A common operating structure in the military is one where a single supervisor oversees several human subordinates for the purpose of managing performance and relaying commands to the appropriate team members. Under information age C^2 structures, the need for this second function will be reduced (even eliminated in some cases), but performance monitoring will still be required. Frequently, these operators will be engaged in HSC tasks, so it will be the job of a supervisor to observe and diagnose HSC issues in one or more teams.

HSC problems can sometimes be subtle in nature, and thus tend to be more difficult to detect than during many other types of operations. Most HSC tasks are primarily cognitive in nature, so the supervisor cannot easily infer accurate performance from physical actions of operators. Rather than being able to directly observe task completion by a human, the supervisor can only evaluate how an operator is interacting with automation that completes that same task, and once it is done, evaluate the results of that effort. Physical actions taken by operators are limited to activities like typing, button pushing, and body movements to position themselves for better screen viewing. Furthermore, the effects of operators’ actions can occur in remote locations from both the supervisor and subordinates. This physical separation means that all people involved with the process must form mental abstractions to envision a complete picture of the situation. Complicating this is that interaction is usually done through artifacts with inherent limitations, such as voice communication, data links, and 2-dimensional screens. While this is clearly a problem with individual operators (it is one of the primary considerations when designing automation of this type), it is an even larger one for supervisors, who must try to synthesize information from multiple operators at once.

Furthermore, isolating a single cause for poor performance of an entire team can be difficult, especially in time-pressured environments characteristic of NCW environments. Lastly, decreases in performance may be the result of automation degradation and have nothing to do with the human. Supervisors may have difficulty separating the two.

The main problem is then how to support supervisors of HSC tasks so that they are better able to understand what their subordinates are doing. Many of the issues previously discussed in this chapter factor into this discussion. In order to quickly observe and diagnose HSC problems, supervisors must have a high level of SA, both for individuals and teams. Even more so than their subordinates, it is critical that HSC supervisors have a clear picture of the team's *overall* situation. The building block to achieving this superior level of SA is access to and absorption of all relevant data. Therefore, information overload will be a particularly acute problem, as a supervisor could be responsible for any or all of the information available to their numerous subordinates. Additionally, due to the greater range of information types received by HSC supervisors as compared to a single operator, the number of possible relationships and behaviors of this data is higher. This means that the complexity of the situation for the supervisor is raised along all three dimensions simultaneously.

Another problematic issue with supervisory monitoring is how to rectify HSC problems once they are detected. There are several options, which may be applied singly or in concert to varying degrees:

1) *Provide a warning to the human whose performance is low.*

It may be the case that operators do not realize that their performance has dropped below acceptable levels, and merely need to be reminded of it and/or motivated to get back on track. An operator's attention may be inappropriately allocated, so a warning provided by the supervisor (who is monitoring their performance) could cue them to correct it. Of course, if an operator is already cognitively overloaded then the warning could have no effect, or even a negative one due to the additional distraction it would provide.

2) *Redistribute the task load between existing team members.*

Workload could be unevenly distributed within a team, or various team members could be more skilled at certain tasks than others at certain times, so dynamically shifting tasks would allow other team members to pick up the slack. If all team members are equally busy or if others lack the expertise needed to perform tasks, this redistribution will not work. Additionally, the complexity of dynamically allocating tasks between team members puts a significant cognitive load on the supervisor.

3) *Change the number of team members or teams on the underperforming task.*

Potential and existing team members or teams could be humans, computers, or a combination of both. It is also important to remember that poor performance can result from under as well as overload. In this case, the supervisor would observe an unacceptable level of performance from a subordinate and initiate a process to either relieve that operator of the most problematic aspects of their tasks or add to their workload, as required. This change could be manifested in one of two ways: 1) A changing of the level of automation experienced by that operator (driven by a supervisor), or 2) Adjusting the individual workload by the addition or subtraction of a human or computer team member. Changing the number of team members requires planning, as the new team members must have access to the correct equipment, programming and training to be effective. As before, there must be an efficient way to reassign jobs between team members. In addition, while changing the number of team members may help in the longer term to reduce overall workload and/or improve performance, there will be a transition period with associated costs for team and individual SA as well as individual and team performance.

4) *Modify the mission objectives or timeline to accommodate lowered overall performance.*

This is a straightforward solution, but not one that will always be available in military situations. Many missions have time-sensitive deadlines that cannot be altered. Similarly, lower level mission objectives may be part of a larger operation, and therefore are not flexible.

Finally, the question of whether the supervisor should be a human or a computer should be discussed. Using a computer offers advantages and disadvantages in this regard. A computer would eliminate information capacity issues and the need for refined HCI designs. However, it would operate on defined rules and would be relatively inflexible in the face of unknown situations. The computer would also lack the capability of a human to predict future performance based on subjective judgments from visual observations and face-to-face interactions.

XX.2.9. Trust and Reliability

A recent survey of Air Force and Air National Guard pilot attitudes regarding the role of UAVs as wingmen for manned aircraft revealed an inherent distrust in highly autonomous systems. Pilots generally thought UAVs were unsuited for a variety of missions to include close air support, search and rescue, and most strike missions. The pilots asserted that only humans are capable of operating in the "free flowing environment" of an offensive combat mission, which requires experience and knowledge to accurately assess the situation and determine a course of action. In addition, pilots did not think that a group of UAVs should be allowed to self-organize. Pilots generally thought UAVs could not make informed decisions about both their individual states and how their capabilities served the current mission. Furthermore, most pilots did not want UAVs operating anywhere near friendly forces on the ground and some did not think UAVs should operate in the same airspace with tactical manned aircraft. Fighter pilots in particular were hesitant to accept a role for UAVs in offensive combat operations, believing that a UAV could never replace a human wingman. An A-10 pilot described his relationship with his wingmen as one of trust and loyalty in that they trained, worked, and fought together, and that a UAV could never replace a human wingman (Cummings & Morales, 2005)

Anecdotal reports from soldiers in Afghanistan using unmanned ground vehicles reveal that soldiers are underutilizing the robots because they inherently distrust the robots, and this inherent distrust in autonomous systems is reflected in the above case study. Distrusting automation when it is perfectly capable has been shown to lead to disuse or misinterpretation of results in order to fit an operator's mental model, even if the subsequent workload caused by the distrust is very demanding and/or time consuming (de Vries, Midden, & Bouwhuis, 2003; Lee & Moray, 1994; Muir, 1987; Parasuraman et al., 1997). In process control, operators' trust in automation has been shown to be primarily based on their perception of the automation's competence (Muir & Moray, 1996), which is likely the case for the soldiers in Afghanistan.

In contrast, other studies have found that pilots tend to trust automation even when it was failing (Conejo & Wickens, 1997; Gempler & Wickens, 1998). As evidenced by the Patriot case study in section XX.2.2 and the discussion on automation bias in section XX.2.6, in time-pressured scenarios with high stakes and uncertainty, operators can trust automated systems too much, to the detriment of the overall mission. Thus designers of NCW systems are faced with a conundrum – how to design a system that is trusted and utilized to its fullest extent, yet not overly trusted such that humans become complacent.

Trust is dynamic in that it changes with exposure to and time between system failures. For example, after an initial system failure, there is a sharp decrease in trust, but it rebounds with consistently correct automation. Subsequently if the automation fails again, trust decreases but it is regained more quickly (Lee & Moray, 1992; Moray, Inagaki, & Itoh, 2000; Muir et al., 1996). Relevant to NCW, recent research in trust and automation reliability in UAV military reconnaissance missions demonstrated that while reliable automation aids operators, unreliable automation can significantly degrade performance (Dixon, Wickens, & Chang, 2004). Specifically, automation that caused a high rate of false alarms for system failures was far more disruptive than automation that failed to alert operators of a failure (otherwise known as a miss). As a result, Dixon et al., (2004) recommended that any automated decision support system that operates below 70% would generate unacceptable costs.

Calibrating an operator's trust a level corresponding to automated system's trustworthiness or reliability is the solution to the problem of too little or too much trust (Moray et al., 2000; Muir, 1987). However, how to do this is still not well understood in the field of human supervisory control. Automation feedback in terms of self-evaluation and interaction with automated decision aids have been suggested as potential strategies for appropriate trust calibration (Lee et al., 1994; Muir, 1987). Displaying the automation's confidence may facilitate better calibration of trust, however, displaying uncertainty and confidence information about an automated recommendation or solution is not a straightforward matter. As discussed previously in section XX.2.6., humans are not intuitive statisticians and tend to introduce biases in probabilistic reasoning (Tversky et al., 1974). Thus presenting probabilistic confidence information to operators so they can make unbiased decisions is not a trivial design problem.

McGuirl and Sarter demonstrated that a categorical trend display of a computer's confidence in its recommendations was superior to a probabilistic static representation for in-flight icing interventions. Uncertainty in information has also been successfully conveyed through degraded images using blended color icons, and expectedly, the addition of numeric probabilities provided no additional advantage (Finger & Bisantz, 2002). Once an operator's trust has been properly calibrated, Xu et al., have demonstrated that even with imperfect automation, human operators can still properly execute their tasking. Given the significant uncertainty that will exist within the actual NCW environment as well as the uncertainty introduced by imperfect automation, more research is needed in trust calibration techniques, especially as they apply to time-pressured decisions.

XX.2.10. Accountability

Figure XX.8 (courtesy of Dr. John Pye of Exponent) illustrates the potentially tragic, albeit unintended, consequences of autonomous system design. The purpose of the UGV (unmanned ground vehicle) in the center of a picture is to enter a hostile area and potentially kill enemy soldiers using its twin double-barreled shotguns. In the scenario in Figure XX.8, the UGV entered a village thought to be inhabited by insurgents, only to be greeted by curious children, obviously fascinated by the seemingly innocent toy. Fortunately the guns were not loaded that day and there were no algorithms in place for automatic firing, but research is currently underway in industry and academia to develop unmanned systems that can independently prosecute targets without human approval.

Figure XX.8: Unintended Consequences of Automation

In addition the myriad of technical issues that surround the NCW human supervisory control problem, there are also social and ethical considerations, especially for weapon systems that impact humans in such a dramatic fashion. What might seem to be the most effective design from a technical viewpoint may not be the most responsible. In one of the few references in the technical literature on humans and automation that considers the relationship between automation and moral responsibility, Sheridan (1996) is wary of individuals "blissfully trusting the technology and abandoning responsibility for one's own actions."

While many technical design issues can be resolved through modeling and testing, degradation of accountability and abandonment of responsibility when using automated systems is a much more difficult question to address. Automated tools are designed to improve decision effectiveness and reduce human error, but they can cause operators to relinquish a sense of responsibility and subsequently accountability because of a perception that the automation is in charge. Sheridan (1983) maintains that even in the information processing role, "individuals using the system may feel that the machine is in complete control, disclaiming personal accountability for any error or performance degradation."

In theory, increased accountability motivates people to employ more self-critical and cognitively complex decision making strategies (Tetlock & Boettger, 1989). In one of the few studies attempting to examine the effects of automation on accountability, Skitka, Mosier, and Burdick (2000) performed an experiment in which subjects were required to justify strategies and outcomes in computerized flight simulation trials. The results showed that not only did increased accountability lead to fewer instances of automation bias through decreased errors of omission and commission, but also improved overall task performance.

How then could systems be designed to promote accountability, especially in the context of NCW? One tangible system architecture consideration for accountability is the number of people required to interact with a given decision support system. Research indicates that responsibility for tasks is diffused when people work in collective groups as opposed to working alone, and this concept is known as "social loafing" (see Karau and Williams, 1993 for a review). This is of particular concern in a distributed system like NCW since task responsibility will often be delegated to many. While research indicates that people experience degraded task responsibility through collective action, the potential loss of a sense of moral responsibility and agency for operators interacting collectively through human-computer interfaces is not as clearly understood. It is likely that the computer interface becomes another entity in the collective group so that responsibility, and hence accountability, can be cognitively offloaded not only to the group, but also to the computer. This is one area in human-computer interaction and accountability research that deserves significantly more attention.

XX.3 SUMMARY

Military forces in the 21st century face complex and subversive threats that often cannot be defeated by conventional tactics. Thus, it is critical that the military be able to leverage all of its available information, and to have sufficient agility to apply the relevant resources to bear on emerging situations. This is the driving force behind the US military's transformation into the Information Age and NCW. However, the primary advantage of operations based upon the tenets of NCW, that of rapid access to information across the network, will likely be a major bottleneck and possible point of failure for those humans who must synthesize voluminous data from the network and execute decisions in real-time, often with high-risk consequences under significant uncertainty. Network-centric operations will bring increases in the number of available information sources, volume of information and operational tempo, all which place higher cognitive demands on operators.

The 10 HSC NCW challenges identified here are not mutually exclusive and apply to other domains such as business and civilian command and control entities such as air traffic control and first response systems. Specifically the adoption of NCW principles will be problematic for human decision makers who need to execute supervisory control across complex, distributed networked systems with a high degree of uncertainty. The implementation of NCW will exponentially add to the number of available information sources as well as the volume of information flow. Without measures to mediate this volume, *information overload* will be problematic. To manage the increase in information across the network, increased *levels of automation* will be needed but often introduce additional human performance problems. One potential design strategy is the use of *adaptive automation*, which has been shown in certain cases to lower workload, but is beset with many technical and mission-critical limitations. Workload mitigation strategies such as increased automation and multimodal displays will increase *complexity*, which can cause a loss of situation awareness or an unexpected and unmanageable increase in mental workload. It is therefore essential that designers be able to measure whether or not interfaces with which NCW operators interact actually reduce complexity instead of add to it.

A more fundamental issue associated with the increase in the number of available information sources, volume of information, and operational tempo under NCW are operator *attention allocation* strategies. NCW hinges on successful information sharing, so knowledge of the relationship between perceived and actual high priority tasks and associated time management strategies, as well as the impact of task disruptions is critical. As a result of NCW information sharing, command and control structures will change significantly. Traditional hierarchical command will be partially replaced by *distributed decision-making* and low-level *team coordination*. Therefore, understanding how to make effective, time-pressured decisions within these organizational structures takes on greater importance in NCW. Moreover, leveraging automation to aid in *supervisory monitoring of operators* is another significant area of concern since NCW will contain embedded HSC systems.

NCW will drive an increase in information-sharing tempo and rapid decision-making. Under these time pressures, the use of heuristics and other naturalistic decision-making methods may be subject to undesirable *decision biases*, both for individuals and groups. Often these decision biases will result in complacent behavior such that operators overly *trust* a complex automated system, but there is also significant distrust of automated systems, which is particularly linked to that system's *reliability*. Lastly, this potentially displaced trust in automation and complacency can lead to a loss of *accountability* and erosion of moral responsibility.

Unfortunately, despite the fact that NCW exists to support human intentions, technological determinism is pervasive in that the primary thrust of NCW research is directed toward improvements and innovations in technology (Alberts, Garstka, & Stein, 2000). The typical but naïve assumption is that advancements in automated systems will naturally improve both system and human performance. Without dedicated focus on how NCW technology impacts both individual and team cognitive processes, as identified in these 10 areas, the DoD vision of greater combat power through the creation of shared situational awareness, increased speed of command, self-synchronization, and higher operational tempo, lethality and survivability will be replaced with a problematic, sub-optimal, and reactive network with significantly increased risk.

XX. 4 ACKNOWLEDGMENTS

This work was an amalgamation of research conducted in the Humans and Automation Laboratory (HAL) at MIT for a number of agencies such as Boeing Phantom Works, the Office of Naval Research, the FAA, and NASA. Special thanks to Michael Biferno of Boeing for motivating the ideas for this work as well as to both the HAL graduate and undergraduate students.

XX. 5 REFERENCES

32nd Army Air and Missile Defense Command. (2003). *Patriot Missile Defense Operations during Operation Iraqi Freedom*. Washington DC: U.S. Army.

Alberts, D. S., Garstka, J. J., & Stein, F. P. (2000). *Network Centric Warfare: Developing and Leveraging Information Superiority* (2nd ed.). Washington, DC: Command and Control Research Program (CCRP).

Alberts, D. S., & Hayes, R. E. (2003). *Power to the Edge*. Washington, DC: Command and Control Research Program (CCRP).

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science, 26*, 39-83.

Altmann, E. M., & Trafton, J. G. (2004). *Task Interruption: Resumption Lag and the Role of Cues.* Paper presented at the 26th Annual Conference of the Cognitive Science Society.

Andreassi, J. L. (2000). *Psychophysiology: Human Behavior & Physiological Response* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Baker, K., Entin, E., See, K., Baker, B. S., Downes-Martin, S., & Cecchetti, J. (2004). *Organizational Structure and Dynamic Information Awareness In Command Teams.* Paper presented at the 2004 International Command and Control Research and Technology Symposium, San Diego.

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Davis, G., Lumicao, M. N., Zivkovic, V. T., Popovic, M. V., & Olmstead, R. (2004). Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired With a Wireless EEG Headset. *International Journal of Human-Computer Interaction, 17*(2), 151-170.

Billings, C. E. (1997). *Aviation Automation: The Search for a Human-Centred Approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bisantz, A. M., Kirlik, A., Gay, P., Phipps, D. A., Walker, N., & Fisk, A. D. (2000). Modeling and Analysis of a Dynamic Judgment Task Using a Lens Model Approach. *IEEE Transactions on Systems, Man, and Cybernetics, 30*(6), 605-616.

Boiney, L. (2005). *Team Decision Making in Time-Sensitive Environments.* Paper presented at the 10th International Command and Control Research and Technology Symposium, McLean, Virginia.

Bolia, R. S., Nelson, W. T., & Vidulich, M. A. (2004). A multi-layer visual display for air battle managers: effects of depth and transparency on performance and workload in a dual-task scenario. *Human Factors and Aerospace Safety, 4*(3), 181-193.

Bolstad, C. A., & Endsley, M. R. (2000). *The Effect of Task Load and Shared Displays on Team Situation Awareness.* Paper presented at the 44th Annual Meeting of the Human Factors and Ergonomics Society.

Carroll, J. M., & Rosson, M. B. (1987). Paradox of the Active User. In J. M. Carroll (Ed.), *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction* (pp. 80-111). Cambridge, MA: MIT Press.

Caterinicchia, D. (2003, June 23). DoD Chat Use Exploded in Iraq. *Federal Computer Week*.

Cellier, J., & Eyrolle, H. (1992). Interference Between Switched Tasks. *Ergonomics, 35*, 25-36.

Conejo, R., & Wickens, C. D. (1997). *The effects of highlighting validity and feature type on air-to-ground target acquisition performance* (ARL-97-11/NAWCONR-97-1). Savoy, IL: Institute of Aviation.

Cooke, N., Kiekel, P. A., Salas, E., Stout, R. J., Bowers, C., & Cannon-Bowers, J. A. (2003). Measuring Team Knowledge: A Window to the Cognitive Underpinnings of Team Performance. *Group Dynamics: Theory, Research and Practice, 7*, 179-199.

Cummings, M. L. (2004a). *Automation Bias in Intelligent Time Critical Decision Support Systems.* Paper presented at the AIAA Intelligent Systems, Chicago, IL.

Cummings, M. L. (2004b). *Human Supervisory Control of Swarming Networks.* Paper presented at the 2nd Annual Swarming: Autonomous Intelligent Networked Systems Conference, Arlington, VA.

Cummings, M. L. (2004c). The Need for Command and Control Instant Message Adaptive Interfaces: Lessons Learned from Tactical Tomahawk Human-in-the-Loop Simulations. *Cyberpsychology and Behavior, 7*(6).

Cummings, M. L., & Bruni, S. (2005). *Collaborative Human-Computer Decision Making in Network Centric Warfare.* Paper presented at the TTCP HUM TP7 Workshop on Aerospace Human Factors Issues in Network-Centric Warfare., Salisbury, UK.

Cummings, M. L., & Guerlain, S. (2004). *An Interactive Decision Support Tool for Real-time In-flight Replanning of Autonomous Vehicles.* Paper presented at the AIAA Unmanned Unlimited, Chicago, IL.

Cummings, M. L., & Morales, D. (2005). UAVs as Tactical Wingmen: Control Methods and Pilots' Perceptions",. *Unmanned Systems, February*.

Cummings, M. L., & Tsonis, C. (2005). *Deconstructing Complexity in Air Traffic Control.* Paper presented at the 49th Annual Meeting of the Human Factors and Ergonomics Society, Orlando, FL.

Cummings, M. L., Tsonis, C., & Cunha, D. C. (2005). *Complexity Mitigation Through Aircraft Structure.* Paper presented at the International Symposium on Aviation Psychology, Oklahoma City.

de Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human-Computer Studies, 58*, 719 -735.

Dekker, A. H. (2002). *C4ISR Architectures, Social Network Analysis and the FINC Methodology: An Experiment in Military Organisational Structure*. Edinburgh, South Australia: Defence Sciences and Technology Organization.

Diethe, T. R., Dickson, B. T., Schmorrow, D., & Raley, C. (2004). *Toward an Augmented Cockpit.* Paper presented at the 2nd Annual Human Performance, Situation Awareness, and Automation Conference, Daytona Beach.

Dixon, S., Wickens, C. D., & Chang, D. (2004). *Unmanned Aerial Vehicle Flight Control: False Alarms Versus Misses.* Paper presented at the Humans Factors and Ergonomics Society 48th Annual Meeting, New Orleans.

DoD. (2001). *Network Centric Warfare: Department of Defense Report to Congress*. Washington DC: Office of the Secretary of Defense.

Donchin, E., Kramer, A. F., & Wickens, C. D. (1986). Applications of Brain Event-Related Potentials to Problems in Engineering Psychology. In S. Porges (Ed.), *Psychophysiology: Systems, Processes, and Applications* (pp. 702-718). Middletown, NJ: Till & Till.

Endsley, M. R. (1988). *Design and Evaluation for Situation Awareness Enhancement.* Paper presented at the Proceedings of the the Human Factors Society 32nd Annual Meeting, Santa Monica, CA.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors, 37*(1), 32-64.

Endsley, M. R., & Jones, W. M. (2001). A Model of Inter- and Intrateam Situation Awareness: Implications for Design, Training, and Measurement. In M. R. Endsley (Ed.), *New Trends in Cooperative Activities: Understanding System Dynamics in Complex Environments*. Santa Monica, CA: Human Factors and Ergonomics Society (HFES).

Fabiani, M., Gratton, G., & Coles, M. G. H. (2000). Event-Related Brain Potentials. In G. G. Berntson (Ed.), *Handbook of Psychophysiology*. Cambridge, England: Cambridge University Press.

Finger, R., & Bisantz, A. M. (2002). Utilizing graphical formats to convey uncertainty in a decisionmaking task. *Theoretical Issues in Ergonomic Science, 1*, 1-25.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Franke, J. L., Daniels, J. J., & McFarlane, D. C. (2002). *Recovering Context After Interruption.* Paper presented at the 24th Annual Meeting of the Cognitive Science Society, Fairfax, VA.

Gempler, K. S., & Wickens, C. D. (1998). *Display of Predictor Reliability on a Cockpit Display of Traffic Information* (ARL-98-6/ROCKWELL-98-1). Savoy, IL: Institute of Aviation.

Gopher, D., Greenshpan, Y., & Armony, L. (1996). *Switching Attention Between Tasks: Exploration of the Components of Executive Control and their Development with Training.* Paper presented at the Human Factors and Ergonomics Society 40th Annual Meeting, Philadelphia, PA.

Hankins, T. C., & Wilson, G. F. (1998). A Comparison of Heart Rate, Eye Activity, EEG and Subjective Measures of Pilot Mental Workload During Flight. *Aviation, Space and Environmental Medicine, 69*(4), 360-367.

Hess, E. H., & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity During Simple Problem Solving. *Science, 132*, 349-350.

Hilburn, B., Jorna, P. G., Byrne, E. A., & Parasuraman, R. (1997). The Effect of Adaptive Air Traffic Control (ATC) Decision Aiding on Controller Mental Workload, *Human-automation Interaction: Research and Practice* (pp. 84-91). Mahwah, NJ: Lawrence Erlbaum.

Humphrey, D. G., & Kramer, A. F. (1994). Toward a Psychophysiological Assessment of Dynamic Changes in Mental Workload. *Human Factors, 36*(1), 3-26.

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

Hutchins, E. (1995b). How a cockpit remembers its speed. *Cognitive Science, 19*, 265-288.

Johansen, R. (1988). *Groupware: Computer Support for Business Teams*. NY: Free Press.

Jones, D. G. (2000). Subjective Measures of Situation Awareness. In D. J. Garland (Ed.), *Situation Awareness Analysis and Measurement* (pp. 113-128). Mahwah, NJ: Lawrence Erlbaum Associates.

Kaber, D. B., & Endsley, M. R. (2004). The Effects of Level of Automation and Adaptive Automation on Human Performance, Situation Awareness and Workload in a Dynamic Control Task. *Theoretical Issues in Ergonomics Science, 5*(2), 113-153.

Kaber, D. B., Endsley, M. R., & Onal, E. (2000). Design of Automation for Telerobots and the Effect on Performance, Operator Situation Awareness and Subjective Workload. *Human Factors and Ergonomics in Manufacturing, 10*(4), 409-430.

Karau, S. J., & Williams, K. D. (1993). Social loafing: a meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*(4), 681-706.

Klein, G. (1989). Recognition-Primed Decisions. *Advances in Man-Machine Research, 5*, 47-92.

Kramer, A. F. (1991). Physiological Metrics of Mental Workload: A Review of Recent Progress. In D. L. Damos (Ed.), *Multiple Task Performance* (pp. 279-328). Washington, DC: Taylor & Francis.

Laughery, K. R., & Corker, K. (1997). Computer Modeling and Simulation of Human/System Performance. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1375-1408). New York, NY: John Wiley & Sons.

Layton, C., Smith, P. J., & McCoy, E. (1994). Design of a cooperative problem-solving system for en-route flight planning: An empirical evaluation. *Human Factors, 36*, 94-119.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics, 35*(10), 1243 - 1270.

Lee, J., & Moray, N. (1994). Trust, self confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40*, 153-184.

Levchuk, G. M., Levchuk, Y. N., Luo, J., Pattipati, K. R., & Kleinman, D. L. (2002). Normative Design of Organizations - Part I: Mission Planning. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 32*(3), 346-359.

Levchuk, G. M., Yu, F., Levchuk, Y., & Pattipati, K. R. (2004). *Networks of Decision-Making and Communicating Agents: A New Methodology for Design and Evaluation of Organizational Strategies and Heterarchical Structures.* Paper presented at the 2004 International Command and Control Research and Technology Symposium, San Diego.

Lord, C. G., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *The Journal of Personality and Social Psychology, 47*, 1231-1243.

Maglio, P. P., & Campbell, C. S. (2000). *Tradeoffs in Displaying Peripheral Information.* Paper presented at the ACM Conference on Human Factors in Computing Systems (CHI 2000).

McFarlane, D. C. (1999). *Coordinating the Interruption of People in Human-Computer Interaction.* Paper presented at the INTERACT'99.

Miller, C. (2000). The Human Factor in Complexity. In J. Weyrauch (Ed.), *Automation, Control, and Complexity: New Developments and Directions*. New York: John Wiley.

Miller, C. A., & Parasuraman, R. (2003, October 13-17). *Beyond Levels of Automation: An Architecture for More Flexible Human-Automation Collaboration.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Denver, CO.

Miller, S. L. (2002). *Window of opportunity: Using the interruption lag to manage disruption in complex tasks.* Paper presented at the 46th Annual Meeting of the Human Factors and Ergonomics Society, Baltimore, MD.

Mitchell, P. M., Cummings, M. L., & Sheridan, T. B. (2005). *Mitigation of Human Supervisory Control Wait Times through Automation Strategies*. Cambridge, MA: MIT Humans and Automation Laboratory.

Miyata, Y., & Norman, D. A. (1986). Psychological Issues in Support of Multiple Activities. In S. W. Draper (Ed.), *User Centered System Design: New Perspectives on Human Computer Interaction* (pp. 265-284). Hillsdale, NJ: Lawrence Erlbaum.

Moray, N., Dessouky, M. I., & Kijowski, B. A. (1991). Strategic Behavior, Workload, and Performance in Task Scheduling. *Human Factors, 33*(6), 607-629.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied, 6*(1), 44 - 58.

Mosier, K. L., & Skitka, L. J. (1996). Human Decision Makers and Automated Decision Aids: Made for Each Other? In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 201-220). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Muir, B. M. (1987). Trust Between Humans and Machines, and the Design of Decision Aids. *International Journal of Man-Machine Studies, 27*(5 & 6), 527-539.

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics, 39*(3), 429-460.

Orden, K. F., Limbert, W., Makeig, S., & Jung, T.-P. (2001). Eye Activity Correlates of Workload during a Visuospatial Memory Task. *Human Factors, 43*(1), 111-121.

Osga, G., Van Orden, K., Campbell, N., Kellmeyer, D., & Lulue, D. (2002). *Design and Evaluation of WarfighterTask Support Methods in a Multi-Modal Watchstation* (1874). San Diego: SPAWAR.

Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., & Barnes, M. (1992). *Theory and Design of Adaptive Automation in Adaptive Systems* (Progress Report No. NAWCADWAR-92033-60). Warminster, PA: Naval Air Warfare Center, Aircraft Division.

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors, 39*(2), 230-253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30*(3), 286-297.

Pew, R. R. (2000). The State of Situation Awareness Measurement: Heading Toward the Next Century. In D. J. Garland (Ed.), *Situation Awareness Analysis and Measurement* (pp. 33-47). Mahwah, NJ: Lawrence Erlbaum Associates.

Polich, J. (1991). P300 in Clinical Applications: Meaning, Method, and Measurement. *American Journal of EEG Technology, 31*, 201-231.

Pope, A. T., Bogart, E. H., & Bartolome, D. (1995). Biocybernetic System Evaluates Indices of Operator Engagement. *Biological Psychology, 40*, 187-196.

Price, J., Miller, D., Entin, E., & Rubineau, B. (2001). *Collaborative Planning and Coordinated Team Performance.* Paper presented at the 6th International Command and Control Research and Technology Symposium, Annapolis, MD.

Prinzel, L. J., III, Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2003). Effects of a Psychophysiological System for Adaptive Automation on Performance, Workload, and the Event-Related Potential P300 Component. *Human Factors, 45*(4), 601-613.

Pritchett, A. R., & Hansman, R. J. (2000). Use of Testable Responses for Performance-Based Measurement of Situation Awareness. In M. R. Endsley & D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 189-209). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Rugg, M. D., & Coles, M. G. H. (Eds.). (1995). *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition*. New York, NY: Oxford University Press.

Scott, S. D., Carpendale, M. S. T., & Habelski, S. (2005). Storage Bins: Mobile Storage for Collaborative Tabletop Displays. *IEEE Computer Graphics & Applications: Special Issue on Large Displays, July/August*.

Shapiro, C., & Varian, H. R. (1999). *Information Rules: A Strategic Guide to the Network Economy*. Boston, MA: Harvard Business School Press.

Sheridan, T. B. (1992). *Telerobotics, Automation and Human Supervisory Control*. Cambridge, MA: The MIT Press.

Sheridan, T. B. (1996). Speculations on Future Relations Between Humans and Automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance* (pp. 449-460). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Sheridan, T. B. (2002). *Humans and Automation: System Design and Research Issues*. Santa Monica, CA: John Wiley & Sons, Inc.

Sheridan, T. B., Vamos, T., & Aida, S. (1983). Adapting Automation to Man, Culture and Society. *Automatica, 19*(6), 605-612.

Sheridan, T. B., & Verplank, W. L. (1978). *Human and Computer Control of Undersea Teleoperators* (Man-Machine Systems Laboratory Report). Cambridge: MIT.

Shilling, R., Morgan, D., Mosbruger, M., Beilstein, D., & Orichel, T. (2003). *Enhancing Information Fusion Using Spatial Auditory Displays and Videogame Interfaces.* Paper presented at the Office of Naval Research (ONR) Workshop on Attention, Perception and Modeling for Complex Displays, Troy, NY.

Simon, P., Rousseau, F., & Angue, J.-C. (1993, October 17-20). *Quantitative Analysis of Mental Workload Influence on Eye Scanning Movements.* Paper presented at the IEEE International Conference on Systems, Man and Cybernetics, Le Touqet, France.

Skitka, L. J., & Mosier, K. L. (2000). Automation Bias and Errors: Are Crews Better Than Individuals? *International Journal of Aviation Psychology, 10*(1), 85-97.

Skitka, L. J., Mosier, K. L., & Burdick, M. D. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies, 51*(5), 991-1006.

Skitka, L. J., Mosier, K. L., & Burdick, M. D. (2000). Accountability and automation bias. *International Journal of Human-Computer Studies, 52*, 701-717.

Smallman, H. S., & St. John, M. (2005). Naïve Realism: Misplaced faith in the utility of realistic displays. *Ergonomics in Design*.

Somervell, J., Srinivasan, R., Vasnaik, O., & Woods. (2001). *Measuring Distraction and Awareness Caused by Graphical and Textual Displays in the Periphery.* Paper presented at the 39th Annual ACM Southeast Conference, Athens, GA.

St. John, M., Cowen, M. B., Smallman, H. S., & Oonk, H. M. (2001). The use of 2-D and 3-D displays for shape understanding vs. relative position. *Human Factors, 43*, 79-98.

St. John, M., Smallman, H. S., & Manes, D. J. (2005). *Recovery from interruptions to a dynamic monitoring task: The beguiling utility of instant replay.* Paper presented at the Human Factors and Ergonomics Society 49th Annual Meeting, Orlando, FL.

Tetlock, P. E., & Boettger, R. (1989). Accountability: A Social Magnifier of the Dilution Effect. *Journal of Personality and Social Psychology, 57*(3), 388-398.

Tulga, M. K., & Sheridan, T. B. (1980). Dynamic Decisions and Work Load in Multitask Supervisory Control. *IEEE Transactions on Systems, Man, and Cybernetics, 10*(5), 217-232.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124-1131.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate. *International Journal of Human-Computer Studies, 57*, 247-262.

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological Workload Reactions to Increasing Levels of Task Difficulty. *Ergonomics, 41*(5), 656-669.

Wells, M., & Hoffman, H. (1998). *Using the Virtual Pilot (ViP) to Enhance Multi-Crew Performance in Complex Military Environments* (R-98-17). Seattle: Human Interface Technology Laboratory.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering Psychology and Human Performance* (3rd ed.). Upper Saddle River, N.J.: Prentice Hall.

Wilson, G. F., & Russell, C. A. (2003). Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks. *Human Factors, 45*(4), 635-643.

Woods, D. D. (1991). The cognitive engineering of problem representations. In J. L. Alty (Ed.), *Human-computer interaction and complex systems*. London: Taylor and Francis.

Yerkes, R. M., & Dodson, J. D. (1908). The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Journal of Comparative Neurology and Psychology, 18*, 459-482.

Zsambok, C. E., Beach, R. B., & Klein, G. (1992). *Literature Review of Analytical and Naturalistic Decision Making*. San Diego: Naval Command, Control and Ocean Surveillance Center.

(Altmann & Trafton, 2002; Boiney, 2005; Cooke et al., 2003; Cummings, 2004a; Cummings & Guerlain, 2004; Franke, Daniels, & McFarlane, 2002; Karau & Williams, 1993; Layton, Smith, & McCoy, 1994; Levchuk et al., 2004; McFarlane, 1999; Price et al., 2001; Skitka, Mosier, & Burdick, 1999; Smallman & St. John, 2005; St. John, Smallman, & Manes, 2005; Tversky, Morrison, & Betrancourt, 2002)