13th ICCRTS

C2 for Complex Endeavors

REVIEW OF COGNITIVE METRICS FOR C2

Topic 6: C2 Assessment Tools and Metrics
Topic 4: Cognitive and Social Issues
Topic 7: Network-Centric Experimentation and Analysis

Mandy Natter, Jennifer Ockerman, and Leigh Baumgart

C4 and Decision Technology Section
National Security Technology Department
Johns Hopkins University Applied Physics Laboratory

Point of Contact:
Mandy Natter
Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel, MD 20723-6099
(240) 228-3994
Mandy.Natter@jhuapl.edu

Abstract

Human cognitive knowledge, skills, and abilities are a significant component of complex command and control (C2), hence measuring cognitive aspects of C2 can provide critical value-added. Cognitive measures provide a consistent gauge to measure C2 cognitive effects. These measures can be used to compare cognitive impacts both between and within systems. Also, these measures help to analyze specific cognitive strengths and weaknesses, so that C2 systems can be improved. Likewise, they can be used to analyze training strengths and weaknesses, and improve training so that it is better suited to user needs.

This paper summarizes an extensive literature review on macrocognitive metrics that apply to complex C2 assessment. Since a suite of cognitive metrics is required to assess C2 warfighters' actual and perceived effectiveness, guidance is provided on selecting appropriate macrocognitive metrics. Mental constructs researched in complex C2 domains including workload, situational awareness, decision making, and collaboration are highlighted. This paper defines each construct, provides measurement tools and techniques, and reviews the costs and benefits of each technique. The paper concludes with an explanation of how the mental constructs and their metrics are inter-related and suggests using several metrics together to assess and explore C2 in complex endeavors.

1 Introduction: Measures overview

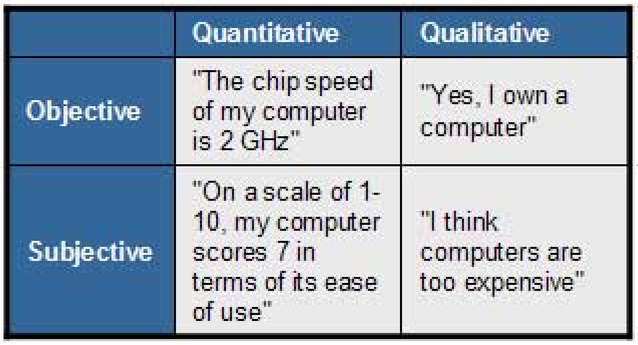
As warfighters in command and control environments are being required to master a larger set of skills for increasingly complex tasks, performance alone is not all that matters in the evaluation or design of a system. The cognitive demands in net-centric operations become increasingly complex as operators must integrate vast amounts of information with varying content, format, age, and degree of uncertainty. Several measure attributes, such as type of measure, measurement scales, number of participants measured, how the measures are rated, who rates the measures, timing, and a variety of other factors affect the measure used.

1.1 Measurement types

Types of measures include quantitative versus qualitative and objective versus subjective. Simply stated, quantitative measures involve quantities (i.e., numbers) and qualitative measures involve written descriptions (i.e., words). Objective measures are often quantitative and are based on exact external measures of things or concepts that exist. Subjective measures are typically qualitative and are based on personal opinion or judgment. However, the pairing of these terms is not exact; Figure 1 shows how these types of measures interact.

1.2 Measurement scales

Four main types of measurement scales exist: nominal, ordinal, interval, and ratio. Nominal scales are distinct and indicate a difference between entities (e.g., A not B). Ordinal scales are lesser or greater relative scales without reference units. They reflect a difference and the direction of difference. Interval scales include equal measurement intervals and do not have end points. Interval scales depict differences, the direction, and magnitude of the difference. Ratio scales are capable of all mathematical manipulations because they have a "true" or defined zero (e.g., length or speed). Common scales used for cognitive metrics include ordinal Likert-type sematic scales (e.g., 1 low and 7 high) or ratio scales (e.g., speed and accuracy) (O'Donnell & Eggemeier, 1986).



	Quantitative	Qualitative
Objective	"The chip speed of my computer is 2 GHz"	"Yes, I own a computer"
Subjective	"On a scale of 1-10, my computer scores 7 in terms of its ease of use"	"I think computers are too expensive"

(Hodgson, 2007)

Figure 1: Interaction between Quantitative, Qualitative, Objective, and Subjective

1.3 Number of participants measured

The number of participants depends on objectives and available resources. If the goal is to become familiar with the data collection process and resources are low, a pilot study can be used. A pilot study is small in scope and contains a small number of individuals. If the goal is to provide a high probability of detecting a significant effect size and the magnitude of the effect (if an effect exists), then a power analysis is required. A pilot study can also be used, and a power analysis conducted, to determine the necessary number of participants to achieve the potential for statistical significance. (Bausell & Li, 2002)

Another aspect to consider in cognitive metrics is if the task is individual or team-based. Most cognitive metrics address the individual, but team measures are needed for C2.

1.4 How measures are analyzed

Typically, descriptive statistics, which describe what the data shows, are run on quantitative data if the distribution is normal. If more resources are available, it might be possible to obtain inferential statistics, which are generalizable to a larger population. Qualitative results often reflect trends and patterns in responses.

1.5 Raters

Raters can include virtually all involved in a study. There are several cognitive metrics that are self reports. Other popular raters are subject matter experts (SMEs), and the experimenter can also rate participants.

1.6 Timing

Cognitive C2 measures are collected either in real-time or post hoc. Real-time is usually preferable, but can be intrusive. Post hoc is less intrusive, but relies on the participants' and raters' memory.

1.7 Measurement Criteria

Given the diversity of techniques available, it is important that the appropriate technique be chosen based on previous research and practical constraints. O'Donnell and Eggemeier (1986) propose that the criteria in Table 1 should be met by any technique to assess workload; however, these criteria can be generalized to any cognitive measure and have been used for non-cognitive measures as well (O'Donnell & Eggemeier, 1986).

Table 1: Measurement Criteria

Criteria	**Description**
Validity	The extent to which the measure is measuring the mental construct of interest. (Zhang & Luximon, 2005)
Repeatability/reliability	The ability to obtain the same results of the mental construct when tests are administered more than once. (Zhang & Luximon, 2005)
Sensitivity	The ability to detect changes in the level of the mental construct imposed by task difficulty or resource demand.
Diagnosticity	The ability to discriminate the amount of the mental construct imposed on different operator resources (e.g., perceptual versus processing versus motor resources).
Selectivity	The ability of a measure to be sensitive to differences only in the cognitive construct of interest (e.g., cognitive demands as opposed to physical workload or emotional stress). (Zhang & Luximon, 2005)
Intrusiveness	The tendency of a technique to interfere with performance on the primary task.
Implementation requirements / Convenience	The ease of implementing a specific assessment technique (e.g., instrumentation requirements or operator training).
Operator Acceptance	The degree of willingness on the part of operators to follow instructions and actually utilize a particular assessment technique.

In order to adequately choose C2 cognitive assessment techniques, it is important to identify the objective of the measurements. Once the objective is identified, it is important to select measurements fitting the objectives with high validity and repeatability/reliability. Next, the sensitivity, diagnosticity, and selectivity can be considered to down-select appropriate measurements. Practical constraints can be further used to screen assessment techniques, with intrusiveness and implementation requirements being more heavily weighted than operator acceptance (O'Donnell & Eggemeier, 1986). In many cases, the best cognitive assessments are derived using multiple techniques in order to capture the complexity of the operators' task demands.

2 Cognitive C2 Metrics

Currently available cognitive metrics relevant to C2 include workload, situation awareness, decision making, and collaboration. Although each of these can be measured in isolation, they are all related to each other. An appropriate workload level should lead to good situation awareness, which in turn should improve decision making and enhance collaboration. This section will provide examples of each of these groups of cognitive metrics. In many cases, as will be noted, more details can be found in the Appendix.

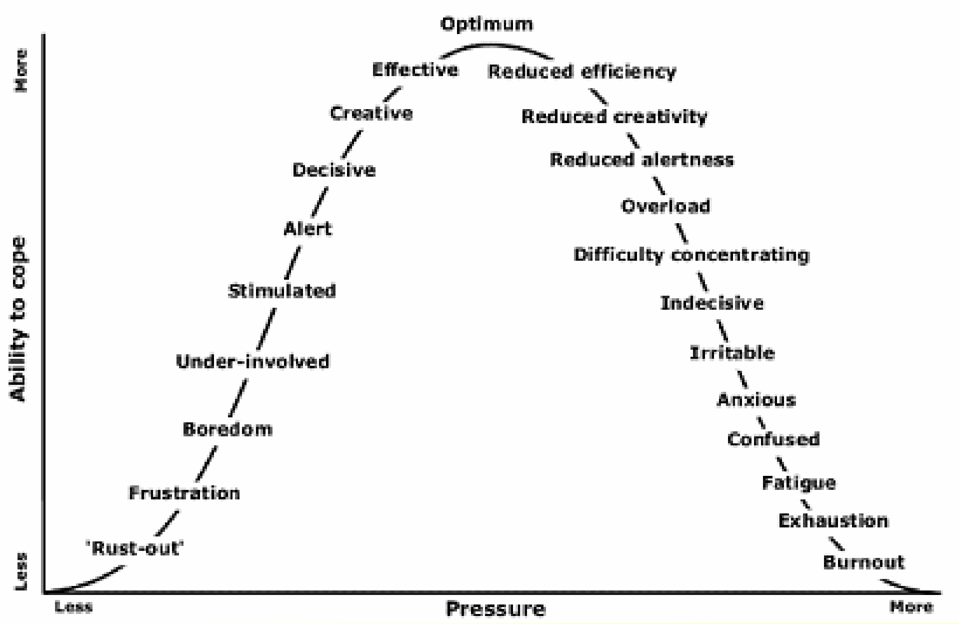
2.1 Workload

The human operator has a limited capacity of resources that can be used to collect, interpret, process, and respond to information in their environment. If the demands of a task exceed an operator's limited capacity, decrements in performance and an increased potential for errors often occur. In the command and control environment, such decrements in performance may be substantial, potentially involving errors that lead to the loss of valuable resources, and even the loss of life.

2.1.1 Defining workload

Workload is often defined as the portion of the operator's limited capacity that is required to perform a particular task (O'Donnell & Eggemeier, 1986). For C2, workload refers to the cognitive demand on the brain and the sensory system (eyes, ears, and skin) due to the task versus physical workload. (Zhang & Luximon, 2005) The assessment of workload can be used to improve either the tools used by an operator, such as enhancing computer interfaces, increasing automation, and reallocating tasks, or the operator by providing more training.

It is important to understand what optimized workload levels are for particular tasks because non-optimal levels of workload may induce stress or boredom. Excess stress often results in changes in information processing, possibly increasing the occurrence of errors. Other consequences of non-optimal workload may include the operator shedding tasks of lower priority, potentially in an unfavorable manner. High levels of workload are not necessarily always "bad." In many environments, low levels of workload coupled with boredom, fatigue, or sleep loss, can have negative impacts on performance as well. The Yerkes-Dodson law demonstrates an empirical relationship between arousal (pressure) and performance (ability to cope) (Figure 2). The law depicts the importance of avoiding extremes of too little or too much workload.



(Haarmann, accessed in 2007)

Figure 2: Yerkes-Dodson Law

2.1.2 Workload assessment techniques

Over the last four decades, a considerable amount of research has been dedicated to understanding and assessing cognitive workload in a variety of domains. Many techniques for workload assessment have been proposed. These techniques generally fall into one of four categories (O'Donnell & Eggemeier, 1986; Wickens & Hollands, 2000): 1) primary-task measures, 2) secondary-task measures, 3) subjective measures, and 4) physiological measures.

2.1.2.1 Primary-task measures of workload

Some researchers have used changes in the quality of operator performance on primary-tasks as a measure of operator workload. It is often assumed that as workload increases, the resources used by an operator will also increase, resulting in a degradation in operator performance. However, this may not always be the case in situations where workload levels increase from very low to moderate and performance may actually improve because the operator becomes less bored and more engaged.

Primary-task workload measures typically include speed and accuracy; however, these measures alone may be insufficient to clearly assess the qualities of the task. For example, good measures of performance on some primary tasks are difficult to define, such as for decision making in the command and control environment. In this environment, the cognitive demand placed on the operator is great, yet the performance outcome result is a function of many other variables aside from the operator's cognitive operations. Further, performance may differ based on data or system limitations, and not on the cognitive demands placed on the operator to perform the task (Wickens & Hollands, 2000).

Using primary-task measures of workload alone also makes it difficult to compare operator workload levels across different tasks or different systems. This is especially apparent when workload is extremely low and achieving perfect performance is easy. In these instances, primary-task measures alone would not be able to distinguish workload differences between the two systems.

2.1.2.2 Secondary-task measures of workload

Secondary tasks are used primarily to simulate realistic workload levels that may not always be experienced in the laboratory or a simulated environment. However, the application of secondary tasks can also be used to assess operator workload. This technique involves imposing a secondary task on an operator in order to measure the residual resources or capacity not utilized in the primary task. Secondary task performance is assumed to be inversely proportional to primary task resource demands. Therefore, differences in primary-task resource demand that are not reflected in primary-task performance alone may be revealed.

When using this technique, the investigator's emphasis should be on the primary task, but the variation in secondary task degradation is measured (Wickens & Hollands, 2000). Operators are instructed to avoid degradations in primary task performance at the expense of the secondary task to ensure that primary task performance is not affected by the secondary task.

Common examples of secondary tasks to measure workload include the random number generation technique, which involves having the operator generate a series of random numbers (Wetherell, 1981). It has been found that the randomness declines as the workload required for primary task increases. Another related technique is to measure an operator's reaction time to certain probes while completing the primary task, as it has been found that reaction time to secondary-task stimulus will increase with increasing primary task workload (Lansman & Hunt, 1982; Wetherell, 1981).

A variant to implementing a secondary task is to use a loading task. In this environment, the operators are asked to devote all necessary resources to the loading task and the degree of intrusion of this task on performance of the primary task is examined. For example, operators could be asked to monitor a gage and push a button each time it falls below a certain level. In this case, degradations in performance for the primary task provide an indication of the resources demanded.

One difficulty to using secondary tasks is that they often interfere with the performance of the primary task of interest. To overcome this, many researchers have employed the use of embedded secondary tasks. These are tasks that are a legitimate component of the operator’s typical responsibilities, but lower in priority, such as an operator responding to a verbal request from a commander as to the latitude and longitude of a friendly object when their main priority is to closely monitor an unfriendly target for movement. Other researchers have used a chat interface to induce information-seeking secondary-tasks during command and control activities (Cummings & Guerlain, 2004).

Another method for implementing an embedded secondary task was demonstrated by Raby and Wickens (1994). In their experiment, secondary tasks were divided into “must,” “should,” and “could” be done. Time spent performing tasks in the three categories were compared between different scenarios of varying workload (defined by time pressure and external communications requirements). This technique provided an accurate means of comparing the differences between workload levels in each of the scenarios. (Raby & Wickens, 1994)

Using secondary tasks to assess workload provides a high degree of face validity as it helps predict the residual resources an operator will have left over in the event of a failure or unexpected event. The same secondary task can also be used to compare the workload of two different primary tasks. However, it is important to consider the different kinds of resources (e.g., vision, hearing, touch) required by a primary task before selecting a secondary task. Workload may be underestimated if the resource demands of the secondary task do not match those of the primary task (Wickens & Hollands, 2000).

2.1.2.3 Subjective measures of workload

Subjective workload assessments elicit the subject’s perception of cognitive loading during a recently completed task. These assessments take the form of questionnaires or structured/unstructured interviews and typically subjects submit self-ratings either manually or through automated systems (Cherri, Nodari, & Toffetti, 2004). Subjective workload techniques include a predefined rating scale that is either one-dimensional or multidimensional with written or verbal descriptions for each level of the scale. One-dimensional scales, which relate one aspect of workload at a time, are beneficial when diagnosticity is important; while multidimensional scales, which deal with several workload factors at one time, are preferable for a global rating of workload more sensitive to manipulations of task demand (Cherri et al., 2004). Because multidimensional scales touch on several factors, when tasks change, ratings are more likely to reflect impacts on one or several of the factors than in one-dimensional scales.

Because of the need for cognitive workload assessment and the limitation of other measures, researchers have proposed several subjective workload assessments (Gawron, 2000). The three most popular techniques are NASA-TLX (Task Load Index), SWAT (Subjective

Workload Assessment Technique), and Modified Cooper-Harper (MCH) Scale, respectively (De Waard, 1996; Rubio, Diaz, Martin, & Puente, 2004; Zhang & Luximon, 2005). NASA-TLX and SWAT are multidimensional scales where subjects weigh the dimensions in the scales by perceived priority, complete the task, and then score the dimensions on a 0-100 bipolar scale, where 0 represents virtually no perceived workload and 100 represents high workload. Selected dimensions can be analyzed separately and/or an overall score can be calculated based on the subject's weights. MCH is a one-dimensional scale where subjects make direct estimates of cognitive loading after completing a task (De Waard, 1996; O'Donnell & Eggemeier, 1986). More details on these methods can be found in the Appendix.

Subjective mental workload methods are popular due to their practical advantages (e.g. low cost, ease of use, and general non-intrusiveness), high face validity, and known sensitivity to workload variations (O'Donnell & Eggemeier, 1986; Zhang & Luximon, 2005). The costs are low as no particular equipment is required. Subjective workload assessments are easy to employ because they are easily accepted and used by subjects. Low primary-task intrusion is secured as long as the scale is administered after completion of the task. Subjective assessments have high face validity because effort is reported directly from the person experiencing the workload. When biases are low, subjective workload methods can be more sensitive than objective measures (Zhang & Luximon, 2005).

The main caveats are confounding factors and short-term memory constraints. One confound is that subjects may confuse different types of task loading (e.g., physical and mental) and personal factors (e.g., mood, fatigue, etc.) with cognitive effort required for the task (O'Donnell & Eggemeier, 1986). Also, not all of the processing done by an individual is available to conscious introspection, therefore hindering the subjective assessment sensitivity. Further, biases such as dislike or unfamiliarity of the task and hesitations to report difficulties affect workload assessments. In addition, because subjective workload assessments are often administered after the task to prevent intrusiveness, subjects may forget elements of effort expenditures (Cherri et al., 2004; O'Donnell & Eggemeier, 1986). Even in instances where subjective assessments are embedded into tasks, they are non-continuous and do not reflect variations in workload during the task.

Subjective assessments should be considered as global indicators of workload versus highly specified diagnostic techniques. Also, it is beneficial to obtain workload ratings as soon as possible after task performance to minimize degradation due to short-term memory limitations.

2.1.2.4 Physiological measures of workload

Various physiological measures have been investigated to objectively measure workload. They include eye measures (e.g., pupil diameter, blink rate and duration, saccade number and duration, and fixation frequency and duration), cardiac measures (e.g., heart rate variability), neural measures (e.g., background brain wave activity and event-related potentials (ERPs)), and skin measures (e.g., electrodermal response (EDR)). There has been limited success with each of them but none of them are currently accurate enough to be used on its own, and a use of several at once is recommended. More details on these measures can be found in the Appendix.

2.2 Situation Awareness

It is critical that human operators have an awareness of what is happening in C2 situations, so that they can understand the tasks they are conducting and the context within which they are working. As mentioned earlier, situation awareness often supports decision making, so improving situation awareness can lead to improved decision making.

2.2.1 Defining situation awareness

In the 1950s, the U.S. Air Force coined the winning element in air-to-air combat engagements in Korea and Vietnam as the "ace factor" or what they called having good situation awareness (SA) (Spick, 1988). Since the term SA originated, it has expanded to include almost any domain that involves humans performing tasks with complex, dynamic systems. As applications have spread and increased, so have SA definitions and measurement techniques. Some SA definitions are human-centric, others are technology-centric, and some encompass the human and the technology, but all generally refer to knowing what is going on and what will happen next. SA is important because it frequently guides decision making and action (Gawron, 2000). The most widely accepted definition is Endsley's human-centric interpretation that "situation awareness is the *perception* of elements in the environment within a volume of time and space (level 1), the *comprehension* of their meaning (level 2), and the *projection* of their status in the near future (level 3)" (Figure 3).

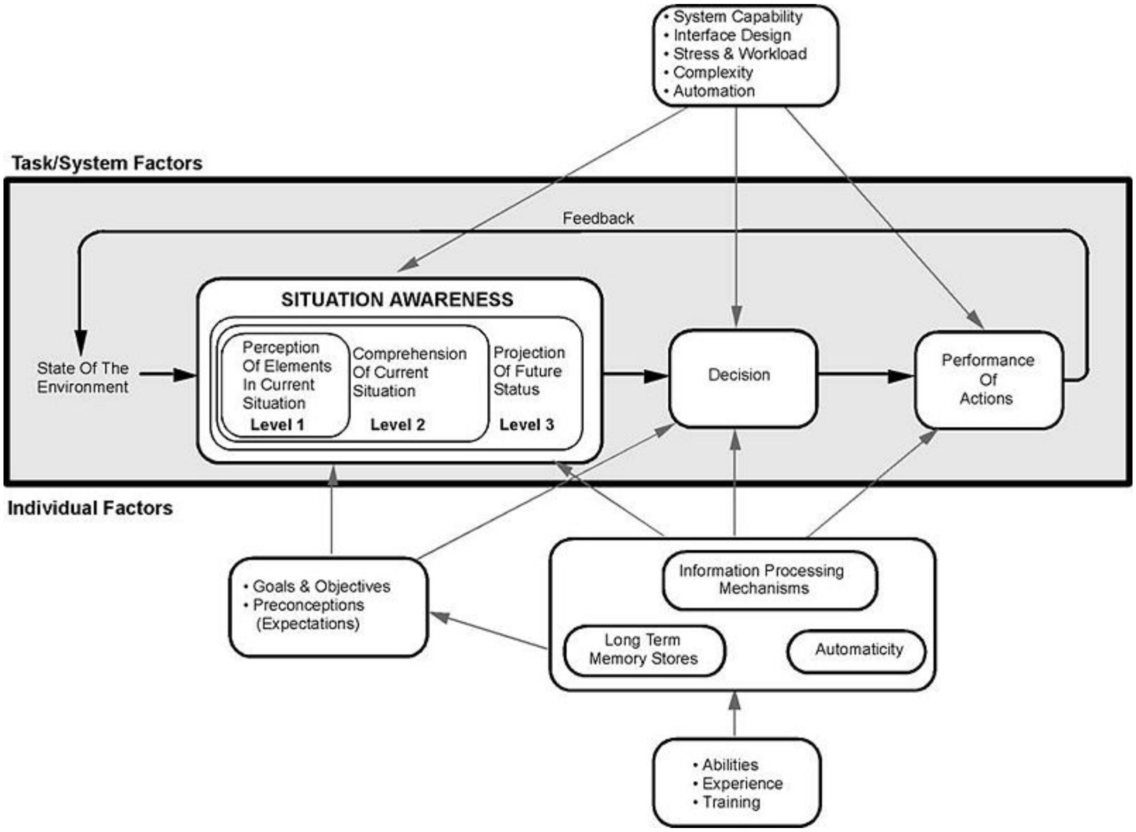


Figure 3: Endsley's Model of SA
(Endsley, 1995)

Military and C2 applications, often call SA situation*al* awareness, which applies to knowledge of the physical elements in the environment (equivalent to Endsley's level 1 SA), while the other levels (equating to levels 2 and 3) are referred to as situational understanding and assessment. (Dostal, 2007) The processes involved with arriving at and maintaining situational understanding in C2 are sometimes called sensemaking. (Gartska & Alberts, 2004)

Technology-centric definitions of SA are linked to C2 applications in that they often refer to the quantity and quality of the information provided by the technology and include data visualization. Technology-centric SA addresses technical challenges of SA, for example, information overload, nonintegrated data, rapidly changing information, high degrees of uncertainty, etc. (Bowman & Kirin, 2006).

Human-system definitions have recently gained maturity and popularity and relate the information provided by the system to the information needed by the operator. Work by Riese et al (2004) and Miller and Shattuck (2004) supply good examples of modeling this relationship. (Miller & Shattuck, 2004; Riese, Kirin, & Peters, 2004) Riese et al.'s model reflects information gleaned from technology (SA_T) being transferred to a human for cognitive situation awareness (SA_C) via an interface (Figure 4). (Riese et al., 2004) Miller and Shattuck's model leverages Endsley's human-centric definition and the lens concept in a multi-step process as shown in Figure 5.(Miller & Shattuck, 2004)The left hand side illustrates the technology part of situation awareness (SA_T) while the right hand side represents the human or cognitive situation awareness (SA_C). As can be seen, some amount of the information from the world is detected by sensors and some amount of that information is made available to the human. The human then perceives the information being displayed, comprehends or makes sense of that information, and finally uses that information to predict what will happen in the world.

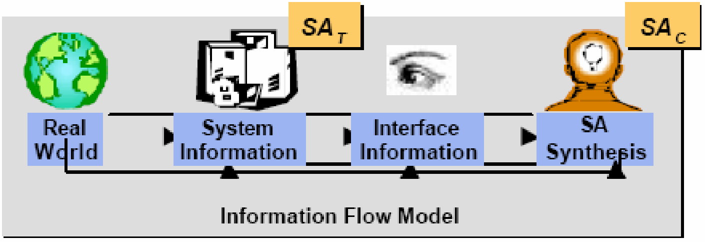


Figure 4: Riese, et al model of SA
(Riese et al., 2004)

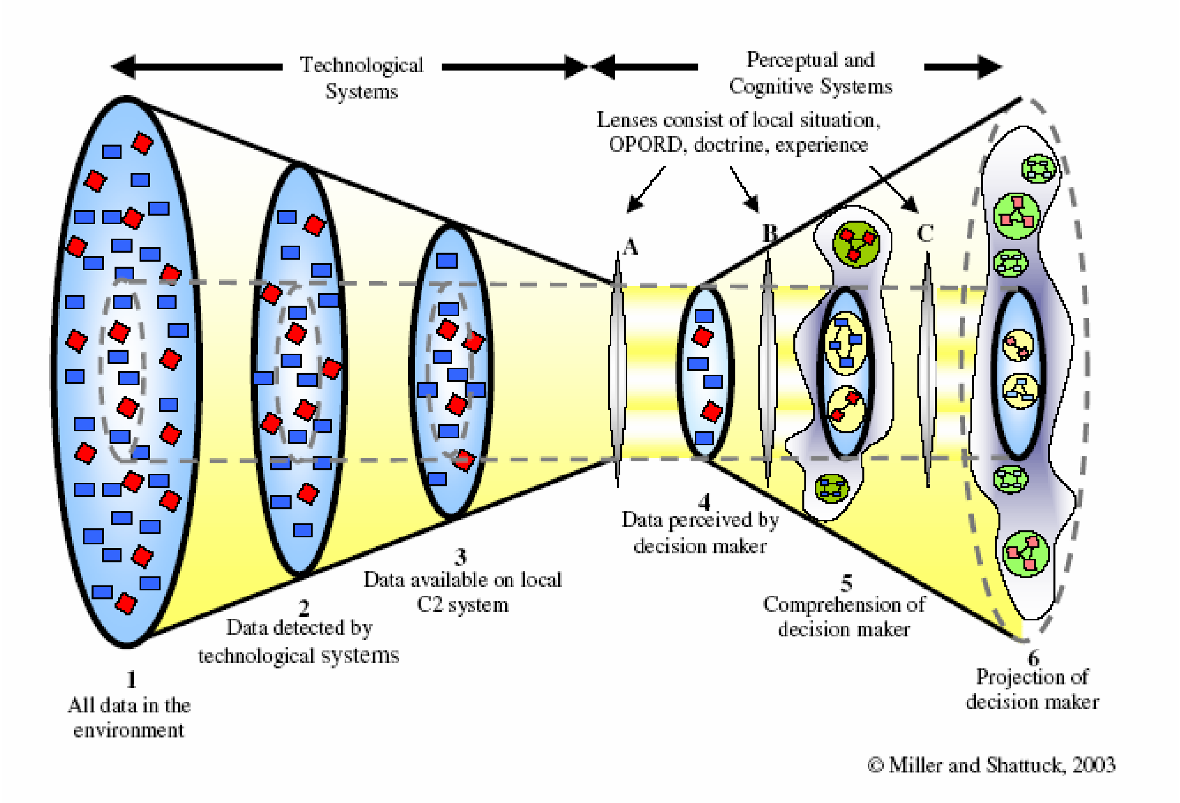

Technological Systems
Perceptual and Cognitive Systems
Lenses consist of local situation, OPORD, doctrine, experience
A
B
C
1
All data in the environment
2
Data detected by technological systems
3
Data available on local C2 system
4
Data perceived by decision maker
5
Comprehension of decision maker
6
Projection of decision maker
© Miller and Shattuck, 2003

Figure 5. A Dynamic Model of Situated Cognition
(Miller & Shattuck, 2004)

2.2.2 Situation awareness assessment techniques

No matter the SA definition utilized, SA is challenging to measure. The information that is required at a particular time, in a particular situation depends on the current goals and objectives of the C2 organization, which are often dynamic. Even when all information is accessible, only a subset of that information is needed to plan and assess the current goals and objectives. Finding the right information at the right time to be aware of what is happening is a challenge, as is leveraging pertinent information to be able to make a decision. In complex, real-world scenarios, it is critical that SA measurement questions and methodology are tailored to the domain and context in which they will be used. There are three main categories of measurement strategies: explicit, implicit, and subjective.

2.2.2.1 Explicit SA measures

Explicit measures assess the users' understanding of what is going on. "Probes," or questions, are administered to prompt subjects to self-report their actual SA. Endsley's three-level information processing definition has a corresponding, validated measurement technique, called SAGAT (Situation Awareness Global Assessment Technique), which is the most commonly used and cited SA metric (Salmon, Stanton, N. Walker, G., & Green, 2006). During SAGAT, the task or simulation is momentarily frozen at various intervals and subjects are asked a set of predetermined multiple-choice questions that relate to the level of SA they have about the situation at that time. Some critics of SAGAT suggest that it measures recall or information decay versus SA; however, Endsley conducted studies reflecting that the freezes do not impact SA probe responses (Gawron, 2000). More recently Endsley's company, SA Technologies, developed a product called The Designer's Situation Awareness Toolkit (DeSAT), which helps to build SA probes and measure SA. Another explicit methodology is to embed opened ended questions throughout the task or simulation, called "real-time probes." This method is less intrusive and more "naturalistic" than interrupting and stopping the task, but due to the open-ended questions, responses can be inconsistent across participants.

2.2.2.2 Implicit SA measures

Implicit measures infer SA from objective, but indirect evidence. In other words, there is an assumption that an objective measure that is not SA implies SA. The difficulty with this logic is that it holds for only simple responses and behaviors because as the complexity increases, the relationship weakens. An advantage of implicit measures is that they are often easier to obtain than explicit measures and are frequently less intrusive. A common implicit SA measure is task performance. For instance, a participant who hits many targets is presumed to have awareness of the targets. In a command and control task, Hiniker 2005 defined SA as the proportion of mission critical warfighter platforms (red, blue, or neutral) correctly identified as important by the commander. The study results revealed that teams with a Common Operating Picture (COP) were able to perform the task 10 percent better than control teams, and therefore had better SA than the control group (Hiniker, 2005). Another implicit SA measure is physiological data (as described in the previous section). Response patterns from some physiological data may correlate with various states or changes in SA state. Communications have also served as an implicit SA measure. For instance, a study by Artman 1999 reflected that the more successful team sent fewer messages between units than the less successful team (Artman, 1999). Another example of an implicit measure is a continuous algorithm developed by Riese, et al to measure instantaneous technical SA or SA_T. SA_T is a ratio of the amount of information provided by the technology/system to the amount of information needed by a decision maker. It is a technique that may be used for a system-centric definition of SA. The study concluded that the relative difference between levels of SA_T was related to battle outcome. In other words, information is linked to force effectiveness. The study also concluded that good SA_T does not necessarily lead to good cognitive SA (SA_C) because humans tend to maintain their beliefs and look for evidence to support their predictions (Riese et al., 2004). More information on calculating SA_T can be found in the Appendix.

2.2.2.3 Subjective SA measures

Subjective measures of SA are the participant's own rating of his/her SA or someone else's ratings of the participant's SA. Some subjective techniques result in a single overall rating (Likert-type scale with a ratings from 1-7), whereas others use multiple scales and break SA into different elements of interest. Subjective methodologies can be fairly generic and administered to practically any operator or decision-maker or be designed to collect specific SA requirements for a particular role or task. Self-ratings, as the name reflects, are the participant's perception of his/her own SA.

The most commonly used self-rating subjective SA technique is called SART (Situation Awareness Rating Technique) developed by Taylor (Endsley, Selcon, Hardiman, & Croft, 1998). SART consists of three bipolar scales: a demand on operator resources, supply of operator resources, and understanding of the situation. These scales are combined, resulting in an overall score. Critics contend that SART may be confounded by performance and workload because it has correlated with both. Proponents highlight that it does not need to be customized for different domains and can be used in both simulated and real-world contexts.

Another subjective measure is observer-ratings, which are specified by someone (usually a subject-matter expert), not involved with the task, observing the subject's performance. Finally, peer-ratings are also used to subjectively evaluate SA. These techniques are employed in teaming experiments where one team member provides a subjective rating of another team member's SA and vice versa. The main advantages of subjective ratings, as noted in the subjective workload measures section, are that they are simple to employ and have high face validity. Also, the experimenter can elicit several ratings during a task for comparison across conditions and time. Subjective ratings also gauge perceived quality or confidence in SA, which can be as important as measuring actual SA, since over-confidence or under-confidence in SA may be just as detrimental as errors in actual SA (Sniezek, 1992). The disadvantages are that they are usually administered after a task and, therefore, rely on memory. Individual differences and inter-rater reliability can confound results (e.g., a rating of a "3" to one person may not be a "3" to someone else). In addition, "unknown unknowns" are an issue, for instance when self-raters overestimate their actual SA because they do not know what they do not know.

2.2.3 Correlation between SA measurement techniques

Studies reveal that explicit and implicit SA measures correlate, but subjective and explicit measures do not correlate. Prince et al. (2007) tested 2-person aircrews on situation awareness using both an explicit (modified SAGAT) and implicit (SMEs administered performance probes) method. Results reflected that both methods could be used to reliably measure SA. Both methods correlated slightly with performance and significantly to each other (Prince, Ellis, Brannic, & Salas, 2007). Endsley et al. (1998) conducted a study comparing explicit (i.e., SAGAT) and subjective (i.e., SART) measurement techniques within a display evaluation. Both SAGAT and SART were sensitive and diagnostic regarding the effects of the display concept. SART highly correlated with subjective measures of confidence; however, SAGAT and SART did not correlate with each other (Endsley et al., 1998).

2.2.4 Team SA measures

Almost all SA measurement techniques refer to individual versus team SA, in part because of the variety of variables and complexity that impacts team SA. For instance, individuals may have degraded SA because they are ignorant of what they are supposed to do, but team SA may be degraded because of status within the team, lack of control, an expectation that another member of the team will take action, actions by team members that prevent an individual's actions, etc. (Prince et al., 2007). Despite this complexity, C2 environments are composed of teams, which are often distributed, and thus require team SA measurement techniques.

Salmon et al. (2006) evaluated seventeen SA measures against a set of human factors criteria for use in command, control, communication, computers and intelligence (C4i) environments. They concluded that current SA measurements are inadequate for use in assessing team SA, and recommended a multiple-measure approach. They identified three requirements for C4i SA measurement: ability to simultaneously measure SA at different locations, ability to measure SA in real-time, and ability to measure both individual and team/shared SA (Salmon et al., 2006).

Measurement techniques for team SA have been researched and continue to be a subject of study (see Appendix for some examples of team SA definitions). Team/shared SA is in a definition phase, and hence explicit measurement methodologies have not been developed. As the definition matures, it is likely that measurement techniques will advance, but in the meantime, multiple-measure approaches combining collaboration and communication via recording how the team responds to unexpected situations is recommended. A recent human-system SA team measurement approach by Hiniker (2005) called the Tech Team Model of SA, includes the Common Operating Picture (COP) as a kind of integral member of the team. In real-world C2 and other stressful environments, implicit measures and subjective measures may be well suited because they are less intrusive than explicit techniques. Regardless of the SA measurement methodology, it is important that it not increase the workload of the participants because increased workload correlates with degraded SA (Entin & Entin, 2000).

2.3 Decision Making

Complex decision making is the most significant human contribution to C2. Some simple decisions can be automated, but often a human is needed to assess risk, weigh alternatives, and select a course of action (COA). Decision making is a critical component of C2 because decision making supports and directs action. If the decision making process can be better and quicker, there is a higher likelihood that C2 will likewise be improved.

2.3.1 Defining decision making

Decision making is a complex process; not just a result. It involves selecting options from alternatives, where some information pertaining to the option is available, the time allotted is longer than a second, and there is uncertainty associated with the selection. (Wickens & Hollands, 2000) An integral component to making a choice is the risk involved with decision execution. The risk, assessed via probability or likelihood, and the consequences of a decision need to be taken into account. (Wickens & Hollands, 2000)

The system (human and technology agents) typically provides the required information, so that decision makers can perceive and interpret that information, generate courses of action (COAs), evaluate the consequences of alternatives, and select a COA from alternatives (Azuma, Daily, & Furmanski, 2006; Endsley, Hoffman, Kaber, & Roth, 2007; Walker et al., 2006). As with SA, decision making can be decomposed into information, or system supported components and cognitive, or human supported, components (Means & Burns, 2005).

The goal of the information component is to present a manageable amount of information to warfighters that pertains to them, but also provides context. Providing the “right” information for decision making depends on having the necessary content with appropriate data attributes. Means and Burns (2005) contend that decisions are information processing tasks, and used data attributes to compare information associated with C2 decisions across three Air Force systems (Means & Burns, 2005). They used Functional Decomposition Diagrams (FDDs) to depict the major goals, decisions, and information in each system. Then, each decision was rated (high or low) on three axes: 1) dimensionality – size of information that must be processed, 2) temporality – a change in the information that must be processed, and 3) uncertainty – ambiguities and probabilities in the information that the decision maker must consider. This technique can help to compare system supported aspects of decision making, and can reveal potential improvements. For instance, when ratings are low across all three dimensions, automation may be implemented; and when ratings are high across all three dimensions, improving visualization may be beneficial to the decision maker (Means & Burns, 2005). The other aspect of identifying the “right” information is the required information content. Information requirements for decision making can be derived through cognitive system engineering knowledge elicitation methods (e.g., task analysis (TA), cognitive task analysis (CTA), cognitive work analysis (CWA), etc.). These methods can reveal the critical fragmentary evidence that experienced decision makers use to create a COA (Azuma et al., 2006).

Content and data attributes may be used to develop a rational model of decision making that can produce a decision autonomously. The rational approach is highly numeric and consists of Bayesian or other probabilistic models of action assessment. Rational models often take all possibilities into account and are advantageous in situations that are not time critical. Since the analyses are numeric, they are amenable to computer logic, but do not map well to how decision makers actually make decisions under time pressure. In addition, rational models frequently do not include a dynamic component, preventing evolution over time. Finally, uncertainty needs to be known a priori to be accounted for in the model. This often leads to compressing uncertainty into probabilities or weightings subjectively assigned, which may obscure the amount, or even presence, of uncertainty. Thus, rational models are difficult in time-critical, uncertain environments like much of C2 execution, but might be beneficial during less time-critical phases such as planning (Azuma et al., 2006).

The decision making strategies employed by humans in C2 more closely match naturalistic decision making (NDM) (e.g., recognition-primed decision making (RPD) (Klein, Calderwood, & Macgregor, 1989). In NDM, the decision maker recognizes the situation based upon features such as: expectancies, plausible goals, relevant cues, and typical action. The solutions are not ideal or optimal, but good enough, especially in time-critical situations. Decisions might be based on fragmentary evidence as it is not feasible to fully quantify the situation and find a mathematic solution, and too much information may be detrimental. If an incorrect decision is made, it can often be changed. The person learns from the decision making process of observation and action. In this way, rational processes are outcome oriented, while naturalistic decision making strategies are process oriented. The disadvantage is the assumption with naturalistic models that the past is a good predictor of the future, which is not always true (Azuma et al., 2006).

Research applying intuitive (NDM) and analytical methods to C2 suggests that a continuum between the two techniques can be used depending on operational context situational resources like computational power, information, and time. For instance, in planning situations, where more time is available, analytic strategies should be used, but in execution, intuitive strategies should be employed because less time is available. Bryant et al (2003) recommend that analytic and intuitive decision making strategies be considered as synergistic styles that when combined can greatly enhance decision making by connecting planning and action (Bryant, Webb, & McCann, 2003).

2.3.2 Decision making assessment techniques

Measuring decision making is complicated because defining a “good” decision is difficult, and many factors and dependencies influence decision making. In order for decision making to be a generalized metric, a good decision should be verified and validated. A method of verifying the decision might be to assume that the optimum decision would produce the maximum value if repeated numerous times. However, this requires defining value, which is often personally subjective and contextually dependent. Also, when repeated, there may not always be a single optimal choice because of differing time and other constraints. Finally, a decision maker may be more concerned with minimizing loss versus maximizing gain. A validation technique might be to assume that good decisions are those that produce “good” outcomes. This is a difficult assumption to maintain in uncertain, probabilistic environments like C2. Adversarial decisions and even own force actions can be unpredictable and impact mission success, which confounds the ability to equate decision making with mission effectiveness (Bolia & Nelson, 2007).

Because decision making is complex, decision making assessments are usually based on what is known and observable. Observing the decision maker can reveal important aspects of decision making, but relies on inferences as to what the decision maker is considering in forming his/her decision. In other words, observations indicate what the decision maker is doing, but not why (i.e., his/her rationale for making the decision, or what major influences led to the decision). Observations or decision products are often used to gauge decisions versus the process because the process is subjective and the products are more objective.

An example of a commonly used observable, objective, result-oriented decision making metric is performance. Decision maker performance is assessed by comparing the decision maker’s decision products to ground truth. Performance is analyzed based on speed and accuracy of the decision(s) versus ground truth. The caveat is that ground truth may not always indicate a “good” or “best” decision. In complex situations like C2, there are several dependencies and risks that influence the outcome, and speed and accuracy can be faulty indicators of “good” decision making. Assuming that the “good” or “best” decision is known, if speed to decision is repeatedly low and accuracy of the decision is repeatedly high, when compared with the ground truth “best” decision, then performance can be a useful decision making assessment metric.

Other objective measures previously referred to in the workload section of this paper and also relevant in decision making are physiological measures. In the Iowa Card Task, where individuals are given four decks of cards and a loan of $2000, participants are told that some decks will lead to gains and some decks will lead to losses. The participants are instructed to use the money as they see fit with the objective to win as much money as possible. Using galvanic skin response measurements, researchers found that individuals micro-sweated more when choosing cards from the disadvantageous decks before they were able to identify which decks were better. “Advantageous” decision making measured through micro-sweating was detected sooner than when participants could verbally explain their card choices. This reflects that cognitive operations are essential to decision making, and emotions and physiological activity may influence decision making (Lamar, 2006). The physiological measures might also be able to speed up the decision making process by identifying the decision before the decision maker is conscious of his/her decision. Also, these physiological measures could potentially be used as performance speed and accuracy decision making measures.

Another observable decision making metric is an expert decision maker’s perception of the decision maker’s ability. This is a subjective technique, but can be useful because the expert can observe the decision maker’s physical processes and deduce whether the decision maker is paying attention to important pieces of information. This technique also reduces the need to have ground truth and a “good” or “best” decision defined because it assumes that the expert will know the “good” or “best” decision. The issue is that experts do not always make “good” or “best” decisions, and they may not always make better decisions than novices. Also, the expert does not know the reason the decision maker made his/her choice.

The issues with these techniques is that they focus more on the decision results than the decision making process. Metrics and measures for decision making should require more than performance results because a correct decision may be the result of chance or luck. "Good decisions may as easily be a fortuitous consequence of ignorance" (Bolia & Nelson, 2007). Even if the individuals making decisions have similar levels of accuracy, they could have very different perceptions of the situation and justifications for their decisions (Cooke, Salas, Cannon-Bowers, & Stout, 2000).

Since decision performance alone is not a sensitive indicator of decision making effectiveness, it may be more appropriate to look at ways that different environmental and informational characteristics influence the processing operations and outcomes of the decision process. First, it would be important to consider how the decision maker gathers and assesses evidence, then it would be beneficial to determine how he/she uses the assessment to make a decision. Information assessment might consist of the sources of information, time to locate the information, and path to locate the information. As mentioned previously, some processing operations that influence decision making are mental constructs like cognitive workload, situation awareness, and sensemaking, so elements of all should be incorporated in decision making assessments (see previous sections of paper) (Bolia & Nelson, 2007).

A number of methods exist to obtain information requirements and decision rationale from the decision maker, the most common being cognitive task analysis (CTA). A variety of CTA methodologies have been developed that differ in approach, structure, emphasis, and resource requirements, but all include some sort of knowledge elicitation, analysis, and knowledge representation (Federal Aviation Administration Human Factors Division, 1999; Militello & Hutton, 1998). (Details on three examples of CTA can be found in the Appendix.) Often the knowledge elicited is not measured quantitatively; however, aspects that can be measured quantitatively are number or type of information/concepts considered and the number or type of consequences considered.

Despite the contrast between CTA techniques, they are useful in revealing C2 decision maker rationale, and all reveal information that could improve C2 processes, or be used to evaluate C2 decision making. The various levels of structure in CTA methodologies parallel the levels of structure in various aspects of C2 (Klein et al., 1989). Another positive aspect is that many of the CTA techniques are conducted retrospectively, which is less intrusive (Klein et al., 1989). The caveat to CTAs are that "no well-established metrics exist" for evaluating CTAs, and it is difficult to evaluate differences between CTA methods (Militello & Hutton, 1998). This is partially because it is unknown what information is lost versus gained in comparison to other techniques and also because interviewees provide different information each time. Also, CTAs can be very resource intensive. Because individual differences impact how much information interviewees are willing to provide and respond, it is difficult to assess the reliability and validity of CTA methods. Finally, no advanced techniques for team CTA have been developed (Militello & Hutton, 1998).

In addition to a lack of verifiable and validated decision making measurement techniques, decision making often involves more than one decision maker or contributor(s) to the decision. This increases the challenge of assessing decision making because, as in situational awareness and workload measures, individual decision making measurement techniques are more mature than team assessments. As described in the next section, the combination of decision making and collaboration assessment techniques are relatively immature.

2.4 Communication and Collaboration

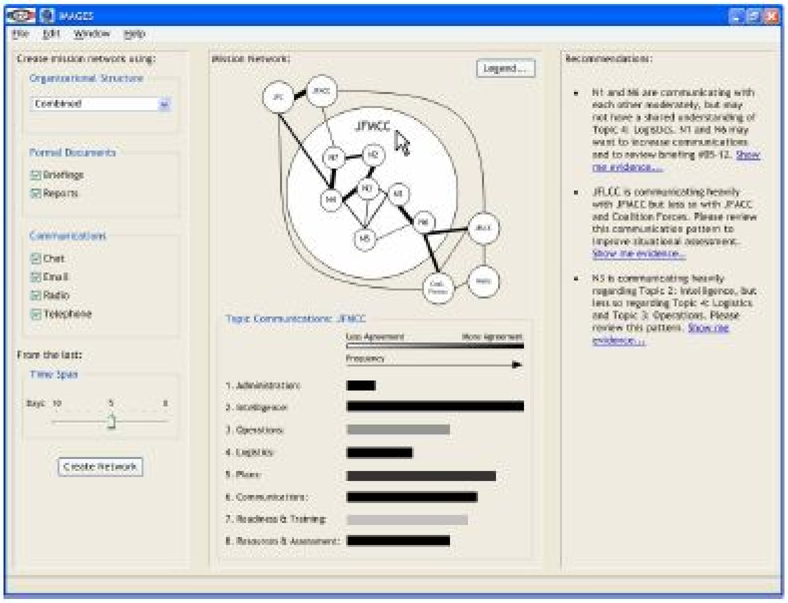
Communication and collaboration are words that are frequently used interchangeably, but important distinctions exist between them. Communication is expression and may include sharing information, but often is from one individual or party to another. Communication is a prerequisite for collaboration, but collaboration involves leveraging the information of others for the purpose of reaching or meeting a goal or objective. It includes synergy of ideas between two or more parties, and depending on how it is defined, can include aspects of situation awareness, workload, and decision making. Both collaboration and communication involve more than one individual, but collaboration often involves a team.

2.4.1 Collaboration assessment techniques

There are a variety of methods used for collaboration assessment. As in other sections of this paper, technological aspects of collaboration will be investigated, and then human aspects of collaboration will be summarized. An example of an assessment of collaboration technology is an analysis of the interconnectivity of team members to information sources, meaning the ability to obtain required information via communication channels. A similar technical support of collaboration that can be assessed is the interconnectivity of team members to each other, in other words whether team members have modes of communication available to collaborate (J. T. Freeman & Serfaty, 2002). In complex C2, often C2 nodes are distributed and collaborate via the Internet (or soon to be GIG), so connectivity issues are important to consider. Net-centric metrics like available bandwidth for communication, communication availability, supported modes of communication, access to communication or collaboration tools, etc. are examples of technological collaboration metrics.

Assessing human aspects of collaboration are usually more complex than technical aspects, depending on the collaboration attribute being analyzed. Relatively simple collaboration metrics to collect and analyze are how much time is spent collaborating, how often various modes of communication are used to collaborate (this may be visualized in a communications usage diagram), and frequency of collaboration. In C2, these simple metrics can be compared across mission phases. Another common collaboration metric used is “who talks to whom,” which is visualized in a form called a social network diagram, consisting of nodes linked by communication lines, often of varying widths to convey frequency or strength (J. Freeman, Weil, & Hess, 2006).

Aptima, Inc. is in the process of testing and refining an automated tool to expand the capability of traditional social network diagrams called the Instrument for Measuring and Advancing Group Environmental Situational awareness (IMAGES) (J. Freeman et al., 2006). IMAGES “captures communications, analyzes them, and presents data concerning ***the distribution of knowledge*** across an organization.” Figure 6 shows a conceptual prototype of some of the functionality (J. Freeman et al., 2006). The mission network in the upper center of Figure 6 is an example of a social network diagram, and the linkages between nodes reflect communication frequency. An important additional attribute of collaboration included in the prototype is collaboration content (in the right panel).



Organizational Structure
Combined

Formal Documents
☑ Briefings
☑ Reports

Communications
☑ Chat
☑ Email
☑ Radio
☑ Telephone

From the last:
Time Span
Days: 10 5 0

- N1 and N6 are communicating with each other moderately, but may not have a shared understanding of Topic 4: Logistics. N1 and N6 may want to increase communications and to review briefing #05-12. Show me evidence...
- JFLCC is communicating heavily with JFMCC but less so with JFACC and Coalition Forces. Please review this communication pattern to improve situational assessment. Show me evidence...
- N3 is communicating heavily regarding Topic 2: Intelligence, but less so regarding Topic 4: Logistics and Topic 3: Operations. Please review this pattern. Show me evidence...

Figure 6: Conceptual prototype of IMAGES

Analyzing collaboration content can be very challenging for researchers because it usually involves manually categorizing collaboration communications by topic, which can be time consuming. Some information that can be obtained through this type of analysis is learning about types of information that require collaboration and collaboration topics. Some other methods to quantify collaboration content include collaboration communications (e.g., number of instances of paraphrasing others), number of critiques initiated concerning high priority issues, and number of gaps, conflicts, and untested assumptions identified (J. T. Freeman & Serfaty, 2002). The IMAGES tool claims to automatically collect collaboration content via Network Text Analysis, which detects the frequency and co-occurrence of terms (J. T. Freeman & Serfaty, 2002). The prototype demonstrates a capability to visualize conflicts amongst topic areas.

Collaboration techniques that analyze content also have the ability to illuminate patterns of collaboration, knowledge, skills and abilities of nodes, and what information is considered relevant or important to various nodes. Another interesting aspect regarding C2 that may be revealed through content analysis is not what is being communicated, but why or the purpose of collaboration. For instance, it may be revealed that the purpose of collaboration is to reach team consensus on a decision. In this scenario, the trade-offs and strategies are communicated and compared, which would help system designers learn about C2 team decision making. Another purpose of collaboration is to reduce uncertainty and this may assist C2 system designers in automation opportunities. Yet another purpose of collaboration may be to achieve a shared understanding or situation awareness or understand the stress and workload of members of the team. All of these issues are critical in complex C2 and require further study so that C2 decision making can be improved and action recommendations can be provided to C2 teams.

3 Applying these Concepts to C2 Research

Unfortunately, there is no silver bullet or simple answer to how to evaluate cognitive aspects of C2 (Barnes & Beevis,). Due to the difficulty of determining what is really occurring within someone’s mind, many of the current measurement techniques are still immature, while many others that are more mature are still contested and debated within the cognitive systems engineering community. A second reason is that even the mature and uncontested measurement techniques are mostly geared towards individuals and there is still a significant amount of research to be done in the area of team evaluation and collaboration. A third reason is the inter-relationships between workload, SA, decision making, and collaboration. These relationships make it difficult to evaluate individual aspects of cognition, as well as, make it undesirable in many cases. This increases the amount of work required to examine cognitive C2 aspects. Finally, many of these measurements cannot be easily collected and analyzed in an automated fashion which makes them very time consuming, labor intensive, and expensive. As if these difficulties were not sufficient, there is also the obstacle of creating environments that can support the evaluation of C2 when access to actual operations is unavailable or too risky.

However, there are ways to mitigate the effects of these difficulties. First, as has been mentioned numerous times throughout this paper, it is best to use a suite of complimentary and overlapping measurement techniques to look at the cognitive aspects from various angles as well as validate other measures. Second, taking the time to carefully design not only the evaluation but also the analysis, with the use of pilot studies and the development of automatic data collection (and analysis), can help reduce the amount of labor needed to fully evaluate a C2 environment. There are many tools available for automatic recording of data from audio and video recording to screen recording and usability suites which allow for correlations between eye, mouse, and screen movements.

This paper has briefly summarized an extensive literature review into currently available cognitive C2 metrics (more details can be found in the Appendix). What is abundantly clear, is that there is still a significant amount of research required to develop reliable, robust, objective, unobtrusive cognitive measurement techniques.

References

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics, 36*, 623-636.

Albers, M. J. (1996). Decision making: A missing facet of effective documentation. *ACM Special Interest Group for Design of Communication: Proceedings of the 14th annual international conference on Systems documentation: Marshaling new technological forces: building a corporate, academic, and user-oriented triangle,* , 57-65.

Allanson, J., & Fairclough, S. H. (2004). A research agenda for physiological computing. *Interacting with Computers, 16*(5), 857-878.

Artman, H. (1999). Situation awareness and co-operation within and between hierarchical units in dynamic decision making. *Ergonomics, 2*(11), 1404-1417.

Azuma, R., Daily, M., & Furmanski, C. (2006). A review of time critical decision making models and human cognitive processes. *Aerospace Conference, 2006 Institute of Electrical and Electronics Engineers, Inc.,*

Barnes, M., & Beevis, D. Chapter 8: Human system measurements and trade-offs in system design. In H. R. Booher (Ed.), *Handbook of human systems integration* (pp. 233-263)

Bass, S. D., & Baldwin, R. O. (2007). A model for managing decision-making information in the GIG-enabled battlespace. *Air and Space Power Journal,* , 100-108.

Bausell, R. B., & Li, Y. (2002). Power analysis for experimental resarch: A practical guide for the biological, medical, and social sciences. New York, NY.: Cambridge University Press.

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovis, M. M., Davis, G., Lumicao, M. N., et al. (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.

Berka, C., Levendowski, D. J., Ramsey, C. K., Davis, G., Lumicao, M. N., Stanney, K., et al. (2005). Evaluation of an EEG-workload model in the aegis simulation environment. Paper presented at the *, 5797* 90-99.

Bolia, R. S., & Nelson, T. (2007). Characterizing team performance in network-centric operations: Philosophical and methodological issues. *Aviation, Space, and Environmental Medicine, 78*(5), B71-B76.

Bowman, E. K., & Kirin, S. (2006). The state of the art and the state of practice: Improving platoon leader situation awareness with unmanned sensor technology.

Bryant, D. J., Webb, R. D. G., & McCann, C. (2003). Synthesizing two approaches to decision making in command and control. *Canadian Military Journal,* , 29-34.

Cherri, C., Nodari, E., & Toffetti, A. (2004). *Information society technologies (IST) programme: Adaptive integrated driver-vehicle interface review of existing tools and methods* No. IST-1-507674-IP)

Collet, C., Petit, C., Champely, S., & Dittmar, A. (2003). Assessing workload through physiological measurements in bus drivers using and automated system during docking. *Human Factors, 45*(4), 539-548.

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R., J. (2000). Measuring team knowledge. *Human Factors, 42*(1), 151-173.

Cummings, M. L., & Guerlain, S. (2004). Using a chat interface as an embedded secondary tasking tool. *Human Performance, Situation Awareness, and Automation Technology II Conference,* Dayton Beach, FL.

De Waard, D. (1996). The measurement of drivers' mental workload. (PhD thesis, University of Groningen).

Dostal, B. C. (2007, Enhancing situational understanding through the employment of unmanned aerial vehicles. army transformation taking shape ...interim brigade combat team newsletter. [Electronic version].(No. 01-18)

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

Endsley, M. R., Hoffman, R., Kaber, D., & Roth, E. (2007). Cognitive engineering and decision making: An overview and future course. *Journal of Cognitive Engineering and Decision Making, 1*(1), 1-21.

Endsley, M. R., Selcon, S. J., Hardiman, T. D., & Croft, D. G. (1998). A comparative analysis of SAGAT and SART for evaluations of situation awareness.

Entin, E. B., & Entin, E. E. (2000). Assessing team situation awareness in simulated military missions. (1) 73-76.

Federal Aviation Administration Human Factors Division. (1999). *Department of defense handbook: Human engineering program process and procedures* No. MIL-HDBK-46855A)

Freeman, J., Weil, S. A., & Hess, K. P. (2006). Measuring, monitoring, and managing knowledge in command and control organizations.

Freeman, J. T., & Serfaty, D. (2002). Team collaboration for command and control: A critical thinking model.

Gartska, J., & Alberts, D. (2004). *Network centric operations conceptual framework version 2.0*

Gawron, V. J. (2000). *Human performance measures handbook*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science, 4*(1-2), 113-131.

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics, 49*(12-13), 1312-1325.

Haarmann, H. *Yerkes-dodson law.*, 2007, from http://www.extra.research.philips.com/probing_experience/presentations/haarmann.ppt#296,29,Yerkes-Dodson-Law

Hiniker, P. J. (2005). Estimating situational awareness parameters for net centric warfare from experiments. *10th International Command and Control Research and Technology Symposium: The Future of C2,* McLean, VA.

Hodgson, P. (2007). *Quantitative and qualitative data - getting it straight.* Retrieved 3/6, 2008, from http://www.blueprintusability.com/topics/articlequantqual.html

Kaempf, G. L., Klein, G., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors, 38*(2), 220-231.

Klein, G. A., Calderwood, R., & Macgregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics, 19*(3), 462-472.

Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems, 21*(4), 70-73.

Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems, 21*(5), 88-92.

Lamar, M. (2006). Neuroscience and decision making.

Lansman, M., & Hunt, E. (1982). Individual differences in secondary task performance. *Memory & Cognition, 10*, 10-24.

Means, C. D., & Burns, K.,J. (2005). Analyzing decisions and characterizing information in C2 systems. *10th International Command and Control Research and Technology Symposium,* McLean, VA.

Militello, L. G., & Hutton, R. J. B. (1998). Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 41*(11), 1618-1641.

Miller, N. L., & Shattuck, L. G. (2004). A process model of situated cognition in military command and control. Paper presented at the San Diego, CA.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. Boff, L. Kaufman & J. Thomas (Eds.), *Handbook of perception and performance* (vol. 2 ed.,). New York: Wiley.

Poythress, M., Russell, C., Siegel, S., Tremoulet, P. D., Craven, P., Berka, C., et al. (2006). Correlation between expected workload and EEG indices of cognitive workload and task engagement. *2nd Annual AugCog International Conference*, San Fransisco, CA. 32-44.

Prince, C., Ellis, E., Brannic, M. T., & Salas, E. (2007). Measurement of team situation awareness in low experience level aviators. *The International Journal of Aviation Psychology, 17*(1), 41-57.

Raby, M., & Wickens, C. D. (1994). Strategic workload management and decision biases in aviation. *The International Journal of Aviation Psychology, 4*(3), 211-240.

Riese, S. R., Kirin, S. J., & Peters, D. (2004). Measuring information availability for situational awareness.

Rowe, D. W., Silbert, J., & Irwin, D. (1998). Heart rate variability: Indicator of user state as an aid to human-computer interaction. Paper presented at CHI ’98 in CHI proceedings, 480-487.

Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology: An International Review, 53*(1), 61-86.

Salmon, P., Stanton, N. Walker, G., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics, 37*, 225-238.

Sniezek, J. (1992). Groups under uncertainty: An examination of confidence in group decision making. *Organizational Behavior and Human Decision Making Processes,* (52), 124-155.

Spick, M. (1988). The ace factor: Air combat and the role of situational awareness. Annapolis, MD: Naval Institute Press.

Van Orden, K. F. (2000). *Real-time workload assessment and management strategies for command and control watchstations: Preliminary findings.* Unpublished manuscript. Retrieved March 16, 2006, from http://www.dtic.mil/matris/sbir/sbir011/Navy89b.doc

Walker, G. H., Gibson, H., Stanton, N. A., Baber, C., Salmon, P. M., & Green, D. (2006). Event analysis of systemic teamwork (EAST): A novel integration of ergonomics methods to analyse C4i activity. *Ergonomics, 49*(12-13), 1345-1369.

Wetherell, A. (1981). The efficacy of some auditory-vocal subsidiary tasks as measures of mental load on male and female drivers. *Ergonomics, 24,* 197-214.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (Third Edition ed.). Upper Saddle River, New Jersey: Prentice-Hall Inc.

Zhang, Y., & Luximon, A. (2005). Subjective mental workload measures. *International Journal of Ergonomics and Human Factors, 27*(3)

APPENDIX

Workload Measurement Techniques

Subjective

NASA TLX, the most commonly used subjective workload scale, contains six dimensions: 1) Mental demand, which refers to perceptual and cognitive activity, 2) Physical demand, which refers to physical activity, 3) Temporal demand, which refers to time pressure, 4) Performance, which is related to personal goal accomplishment, 5) Effort, which refers to energy expenditure in accomplishing the required level of performance, and 6) Frustration, which is related to feelings of irritation, stress, etc. Subjects complete a pair-wise comparison procedure to weight the dimensions before the task. After completing the task, subjects rate the six dimensions using a 0-100 scale (Hart & Staveland, 1988).

The SWAT contains three dimensions: 1) Time (T) load, which reflects the amount of spare time available in planning, executing, monitoring a task; 2) Mental effort (E) load, which assesses how much conscious mental effort and planning are required to perform a task; and 3) Psychological stress (S) load, which measures the amounts of risk, confusion, frustration, and anxiety associated with task performance. The three SWAT dimensions (i.e. T, E, and S) are at three discrete levels (i.e. 1, 2, and 3). Subjects rank the three dimensions and three levels by perceived importance in a 3^3 or 27-card sorting exercise before conducting the task. The card sort results in seven weighting schemes: TES, ETS, SET, TSE, EST, STE, and equal emphasis on T, E, and S. After completing the task, subjects rate the task on the three dimensions (Reid & Nygren, 1988).

Wierwille and Casali (1983) modified the wording of the validated physically focused Cooper-Harper Rating Scale such that it would be appropriate for assessing cognitive functions like “perception, monitoring, evaluation, communications, and problem solving.” The MCH scale maintains a decision tree architecture where participants respond to yes or no questions that lead to options for rating. The rating scale is 1-10, where 1 is very easy and 10 is impossible (O'Donnell & Eggemeier, 1986).

Physiological

Eye Measures

A variety of studies have shown that various aspects of eye behavior correlate with cognitive workload. One of the most sensitive eye physiological measures is pupil diameter. Pupil diameter increases (dilates) as cognitive workload increases (Ahlstrom & Friedman-Berg, 2006). Pupil diameter changes can be dynamic, for instance during comprehension of individual sentences, or sustained during recall of digit span (Van Orden, 2000). Although the average pupil diameter changes by as much as 0.6mm when recalling seven digits, many confounds such as ambient lighting, stimulus characteristics, and even emotional effects can cause pupillary responses that are greater than those from workload alone (O'Donnell & Eggemeier, 1986; Van Orden, 2000). Also, accurate measurement techniques required may impose constraints on experimentation by requiring the subject to stay in one location or wear a measuring device on his/her head. Pupil diameter measurements are therefore difficult to use in applied settings where the environment and other external factors are not controlled. Finally, research suggests that pupil diameter measurements may be highly responsive to cognitive workload changes, yet not diagnostic because there is little ability to identify the resource (e.g. visual, auditory, etc.) utilized in the task (O'Donnell & Eggemeier, 1986).

Generally, blink rate and blink duration decrease as workload increases (Ahlstrom & Friedman-Berg, 2006; Van Orden, 2000; Veltman & Gaillard, 1998). Boehm-Davis et al. (2000) suggest that eye blinks are suppressed when individuals are engaged in cognitive processing; however, eye blinks show great variability (Boehm-Davis, Gray, & Schoelles, 2000; O'Donnell & Eggemeier, 1986). Blink duration is also somewhat unreliable to gauge cognitive workload because other factors like visual workload confound cognitive workload. A visual tracking task with minimal cognitive load can cause lower blink durations than during a more cognitively challenging flight simulation task (Van Orden, 2000). Therefore, eye blinks and blink duration should be considered global indicators of long-term effects versus specified diagnostic techniques (O'Donnell & Eggemeier, 1986).

Another potential eye related measurement of workload is number and duration of saccades. The number of saccades, which are a series of small, quick, jerky movements of the eyes when changing focus from one point to another in the visual field, increase as workload increases. Saccade duration, which typically lasts for 20 to 35 milliseconds, decreases as workload increases. (De Waard, 1996; Poole, 2004; Wickens, Mavor, & McGee, 1997) While saccades may provide clues about the cognitive strategy employed, studies reflect that prior to voluntary eye movement, attention shifts to the location of interest; therefore saccades may be measures of attention versus cognitive workload (Tsai, Viirre, Strychacz, Chase, & Jung, 2007).

After each saccade, the eyes stay still and encode information in movements called “fixations.” Fixation frequency and fixation duration or dwell time both increase as cognitive workload increases, but is task dependent (De Waard, 1996; Poole, 2004; Van Orden, 2000). For example, during a challenging flight simulation, fixation duration correlated with the number of flight rule errors, reflecting a correlation with cognitive workload; however, in a challenging visual search task, search fixation frequency increased and fixation duration did not change (Van Orden, 2000). Fixation is particularly sensitive to visual workload, making it more diagnostic than other techniques; however, fixation does not necessarily imply cognition (De Waard, 1996).

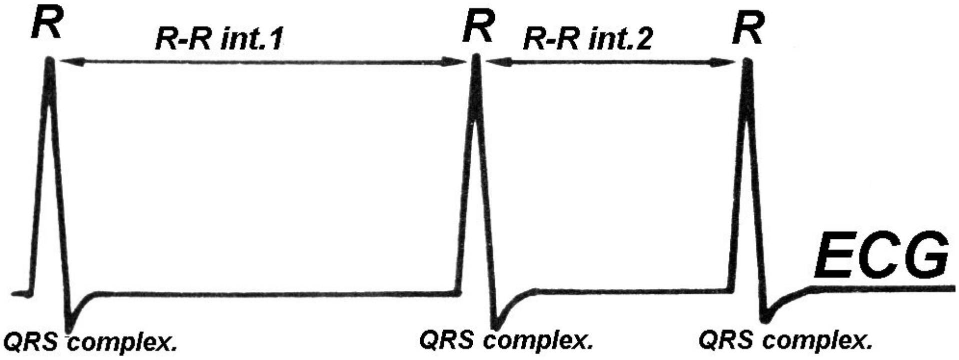
Scan paths, recurring patterns of saccades and fixations, become less of a pattern between display elements as workload increases (O'Donnell & Eggemeier, 1986; Poole, 2004). Also, dwell times in each position lengthen and fewer display elements are used. Scanning is typically a global indicator of workload, but scanning may be a diagnostic index of the source of workload within a multi-element display environment if (1) critical information must be gathered from multiple locations, (2) relative importance of data obtained from each location is different, and (3) the subject can adjust or change the imposed load by a change in strategy (O'Donnell & Eggemeier, 1986).

In summary, blink rate, blink duration, and saccade duration all decrease while pupil diameter, the number of saccades, and the frequency of long fixations all increase with increased workload.

The eye is readily accessible to observation and provides rich data that can be assessed, but study results differ in eye physiological measures that correlate highest with cognitive workload. For example, in a mock anti-air warfare task, blink frequency, fixation frequency, and pupil diameter were the most predictive variables correlating eye activity to target density (Van Orden, 2000); and in an air traffic controller study, managing traffic during adverse weather conditions, decreased saccade distance, blink duration, and pupil diameter correlated closest to cognitive workload (Ahlstrom & Friedman-Berg, 2006). Because of the variability amongst eye physiological measures, it is recommended that multiple techniques be combined (Van Orden, Limbert, Makeig, & Jung, 2001). Several confounds including visual workload impact results, so eye measures should be used as global indicators of cognitive workload. Eye physiological measures provide more sensitive results in controlled environments.

Cardiac Measures

The main cardiac measures studied for sensitivity to cognitive workload include the electrocardiogram, blood pressure, and blood volume. Of these three, measures of electrocardiographic activity show the most promise (Rowe, Silbert, & Irwin, 1998). The electrocardiograph produces a graphic called an electrocardiogram, abbreviated ECG or EKG, from the German *elektrokardiogramm*, which records the electrical activity of the heart over time. Surface electrodes are placed on the skin of a subject to identify pulse beats, recognizable by a pattern called the QRS complex (Figure 1) (O'Donnell & Eggemeier, 1986).



(Dantest Medical System,)

Figure 1: Heart Rate and Heart Rate Variability

Although absolute heart rate has been used as a measure of overall workload, spectral analysis of heart rate variability (HRV) or sinus arrhythmia reflects some correlation to cognitive workload (De Waard, 1996; O'Donnell & Eggemeier, 1986; Tattersall & Hocky, 1995; Veltman & Gaillard, 1998). As the name implies, HRV is the variability in heart rate or the variability between R-R intervals (see Figure 1). Of the over 30 techniques available for determining HRV (e.g. Fourier transform, autoregressive modeling, time-varying analysis, broadband spectral analysis, etc.), most cognitive loading HRV measures emphasize frequency. Frequency HRV techniques measure the amount of variation in different frequency bands. There are three major frequency bands: (1) very low-frequency band (0.0033-0.04 Hz), associated with temperature regulation and physical activity, (2) low-frequency band (0.04-0.15 Hz), associated with short-term regulation of arterial pressure; and (3) high-frequency band (0.15-0.40 Hz), reflecting the influence of respiration. Several studies have suggested that the low-frequency band, and specifically what is called the 0.10 Hz component, indicates cognitive workload (De Waard, 1996; Nickel & Nachreiner, 2003; O'Donnell & Eggemeier, 1986). The 0.10 Hz component reflects short-term changes in blood pressure. A peak of the 0.10 Hz component reflects decreased cognitive workload, and a flattening of the 0.10 Hz component reflects conditions of greater mental workload (Rowe et al., 1998).

Nickel and Nachreiner (2003) assessed the diagnosticity (i.e., the ability to differentiate amongst different types of tasks) and sensitivity (i.e., the ability to detect levels of difficulty) of the 0.1 Hz component of heart rate variability (HRV) for cognitive workload using 14 cognitive tasks (e.g. reaction time, mathematical processing, memory-search, grammatical reasoning task, etc.) from an environmental stressors standardized test in a laboratory context. Only one type of task could be discriminated as different from the other types of tasks – that task reflected a cognitive loading score that matched the cognitive loading expected at rest; however, these results directly conflicted with performance (i.e., performance errors were made) and perceived difficulty (i.e., participants reported mental workload). In terms of sensitivity, the results echoed several other studies that HRV can discern between work and rest, but not to gradations in between (Rowe et al., 1998). Because the experimenters noted differences in the 0.10 Hz component when time pressure was involved, they propose that HRV be used as an indicator for emotional strain or time pressure versus cognitive workload (Nickel & Nachreiner, 2003).

Research by Hannula et al. (2007) supports the use of HRV as a stress indicator. They applied an artificial neural network analysis to evaluate the relationship between cognitive workload that raises the psychophysiological stress and HRV data in fighter pilots. The Pearson's coefficients between the ECG data and the cognitive workload that increases psychophysiological stress as evaluated by an experienced flight instructor were between 0.66 and 0.69 (Hannula, Koskelo, Huttenen, Sorri, & Leino, 2007).

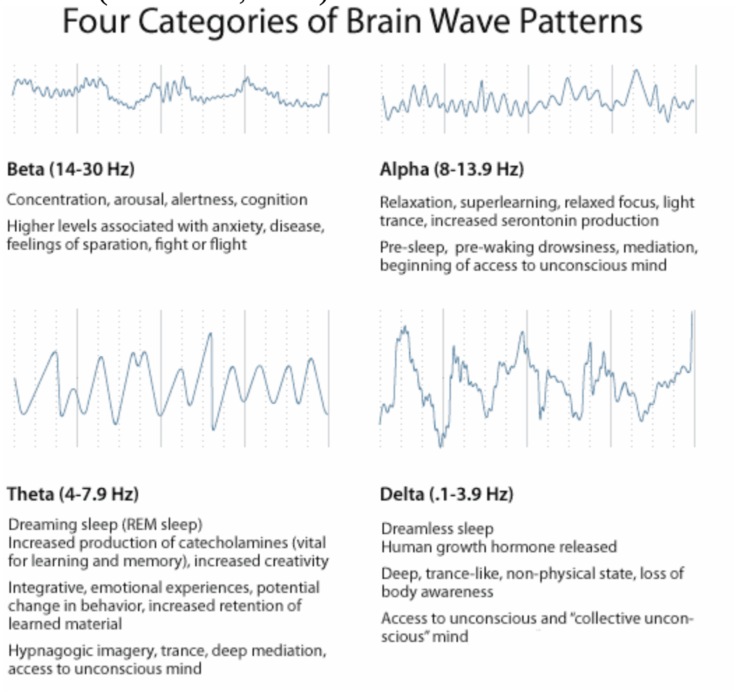
There are several caveats to using heart rate variability as a cognitive workload measure. Possibly the most significant disadvantage to HRV is that it has not been validated as a sensitive cognitive workload indicator (O'Donnell & Eggemeier, 1986). Also, heart rate and, likewise to a smaller extent, HRV are confounded by psychological processes like high responsibility or fear, physical effort and speech, and environmental factors such as high G-forces (De Waard, 1996; O'Donnell & Eggemeier, 1986; Tattersall & Hocky, 1995). As indicated in the studies discussed above, HRV is influenced by stress and time constraints. Physical effort will impact HRV results unless it is kept to a minimum and constant across conditions (De Waard, 1996). Speech can also confound HRV results if verbalization is longer than 10 s and relatively frequent (more than one to five times per minute) (De Waard, 1996). Another factor that affects HR measures, and to a lesser degree cognitive workload, is age. If HRV is the primary workload measure, it may be necessary to restrict elderly subjects from participation because HRV may decrease with increasing age. Finally, a last caveat to consider is that operators typically need to act as their own control because of the idiosyncrasies in the measure (De Waard, 1996).

Despite the caveats, heart rate measurement is arguably the simplest physiological index to measure and it has been employed extensively. The ECG signal requires minimal amplifying (approximately 10 to 20 times less than continuous EEG) and if measurements are limited to R-wave detection and registration, then electrode placement is not critical (De Waard, 1996). Cardiac techniques are also the most popular physiological technique in the last 40 years (Rowe et al., 1998). They are relatively noninvasive and unobtrusive (O'Donnell & Eggemeier, 1986). Also, with continuously recorded cardiac measures, research has shown that HRV can indicate within seconds the change from work to rest (Rowe et al., 1998).

Neural Measures

Electroencephalography, the measurement of electrical activity in the brain, is the most common neurophysiological technique used as an indicator of cognitive workload (Berka et al., 2004). It typically involves a noninvasive procedure of placing electrodes on the surface of the head to detect activity through the skull and scalp. In rare instances, electrodes are placed subdurally or in the cerebral cortex. The traces of activity that result are called an electroencephalogram (EEG), which represents the electrical signals or postsynaptic potentials from a large number of neurons. In clinical use the EEG is considered a “gross correlate of brain activity” because instead of measuring electrical currents from individual neurons, it reflects relative voltage differences amongst various brain areas.

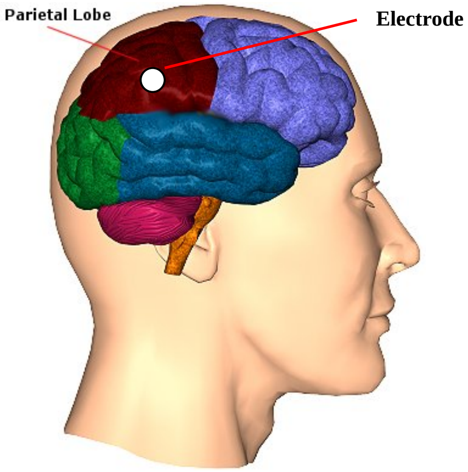
Frequency analyses performed on EEG signals are also called epoch analyses or background EEG analyses and usually result in four ranges or wave patterns (Figure 3) (De Waard, 1996). Although these wave patterns or bands are traditionally used to provide information about the health and function of the brain, some are very responsive to variations in alertness and attention. Specifically, several studies confirm that alpha and especially theta bands are sensitive to aspects of short-term or working memory (Gevins & Smith, 2003). When working memory is in use, research reflects decreases in the upper alpha band and increases in the theta band that become more exaggerated when load increases. This suggests that these bands may be indicators of cognitive loading. One study reflected decreased alpha and increased theta activity during dual-task performance when compared to single-task performance. However, individual differences may be significant; for example, a small number of individuals do not generate alpha waves at all (De Waard, 1996).


Four Categories of Brain Wave Patterns
Beta (14-30 Hz)
Concentration, arousal, alertness, cognition
Higher levels associated with anxiety, disease, feelings of sparation, fight or flight
Alpha (8-13.9 Hz)
Relaxation, superlearning, relaxed focus, light trance, increased serontonin production
Pre-sleep, pre-waking drowsiness, mediation, beginning of access to unconscious mind
Theta (4-7.9 Hz)
Dreaming sleep (REM sleep)
Increased production of catecholamines (vital for learning and memory), increased creativity
Integrative, emotional experiences, potential change in behavior, increased retention of learned material
Hypnagogic imagery, trance, deep mediation, access to unconscious mind
Delta (.1-3.9 Hz)
Dreamless sleep
Human growth hormone released
Deep, trance-like, non-physical state, loss of body awareness
Access to unconscious and “collective unconscious” mind

(Harris,)

Figure 3: Four Categories of Brain Wave Patterns

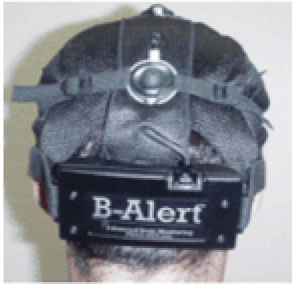
In addition to background EEG, event-related potentials (ERPs) are a neural measurement technique used as a cognitive workload indicator (Berka et al., 2004). Unlike evoked potentials, which are the result of physical stimuli, ERPs may be caused by processes such as memory, attention, expectation, or changes in mental state(Basar-Eroglu & Demiralp, Jan 2001). ERPs are any response to an internal or external stimulus, but are often considered to be the direct result of cognitive activity. ERPs can be reliably measured from voltage deflections in EEG. For instance approximately 300 milliseconds following unpredictable stimuli, there are positive deflections in voltage called the P300 or more simply, P3. The P3 peaks around 300 ms for very simple decisions, and the amplitude increases with unexpected, task-relevant stimuli. After 300 ms, the latency of P3 increases with task difficulty and relates to cognitive workload (De Waard, 1996). During primary tasks only, the P3 amplitude increases with task complexity; however when secondary tasks are added, the P3 decreases with primary task complexity (De Waard, 1996). An external factor is age because the P3 latency increases between 1 and 2 milliseconds each year of the adult lifespan (Colman, 2001). P3s are somewhat difficult to detect, and therefore it is often necessary to present an individual stimulus dozens or hundreds of times and average the results together to cancel out the noise and present the stimulus response clearly. The signal to noise ratio can be improved by placing electrodes on the participant's head above the parietal lobe of the brain (Figure 4).



(Department of Kinesiology University of Waterloo,)

Figure 4: Parietal sites for electrodes

The B-Alert System (Figure 5), a device developed by Advanced Brain Monitoring (ABM), is a commercially available product that claims to have a cognitive workload measurement capability. Originally created to provide early detection of drowsiness, B-Alert advertises real-time workload calculations via a wireless, light weight EEG headset that can analyze six channels of EEG. The signal processing including amplification, digitization, and radio frequency transmission of the signals, is built into a portable unit worn with the headset (Berka et al., 2004).



(Berka et al., 2005)

Figure 5: B-Alert System

Berka, the president of the B-Alert system, et al. (2005) conducted an experiment to determine if EEG could be used as an indicator of cognitive workload with the B-Alert System. The context of the experiment was a simulated Aegis command and control environment, a combat system with advanced, automatic detect-and-track, multi-function phased array radar. Five participants were trained as identification supervisors (IDSs). IDSs have one of the highest workloads of the 32 operators in a typical Aegis Combat Information Center (CIC). The IDSs were responsible for monitoring multiple data sources, detecting required actions, responding appropriately, and maintaining system status within predefined desirable parameters. Workload measures were calculated in real-time for each second of EEG by the B-Alert system. The B-Alert system used signal analysis techniques to decontaminate eye blinks, EMG data, amplifier saturation, and excursions related to movements. Post-hoc analysis identified the cognitive tasks from most difficult to least difficult were track selection-identification, alert-responses, hooking-tracks, and queries. High/extreme workload was detected approximately 100% of the time during high cognitive-load tasks such as selection-identification and alert-responses, 77% of the time for hooking-tracks, and 70% of the time for queries. The low false alarm rate of < 5% reflected that the workload gauge was not overly sensitive (Berka et al., 2005).

Although these results seem encouraging, currently no other known authors have reported such significant success with the B-Alert system and with the signal analysis techniques. A usability evaluation of a prototype of the Tactical Tomahawk Weapons Control System with the B-Alert system reflected that EEG correlated with expected workload when EEG measures were averaged across *all* nine participants. Individual participant EEG measures did not correlate strongly with cognitive workload, reflecting that the EEG was not a very sensitive indicator of cognitive workload. It was also found that EEG did not correlate with expert subjective ratings, but this may be due to insufficient data (Poythress et al., 2006).

The main benefits of EEG measurements are that they are continuous, cognitively unobtrusive, sensitive, moderately diagnostic, and relatively inexpensive. EEG provides a relatively continuous stream of data that can be identified and quantified on a second-by-second basis. Also, the decrease in size of electrodes and the development of wireless systems like B-Alert have led to minimal interference in primary task performance, unlike other functional neuroimaging techniques that require the subject to be completely immobile and are much more massive (Gevins & Smith, 2003). EEG is sensitive to changes in task complexity and task difficulty (Berka et al., 2004). The sensitivity of EEG is high and somewhat diagnostic as it can reflect subtle changes in attention, alertness, and cognitive workload (Berka et al., 2005; Gevins & Smith, 2003). Also, the technology required for EEG is fairly low in cost (Gevins & Smith, 2003).

However, the cost of the sensitivity is a poor signal-to-noise ratio. EEG can be easily contaminated by electrical signals of the eyes, muscles, heart, and external sources (De Waard, 1996). Also the considerable intra-subject and between subject variability makes it difficult to find consistent patterns of physiological change (De Waard, 1996). Signal analysis techniques prevent some of the signal-to-noise ratio problems; however, they must be rigorous and are time consuming. Required software and hardware is costly, and wired electrode systems may physically constrain participants. Advances have been made, but there is still a great deal of work to be conducted to accurately apply EEG measurements to cognitive workload assessment (Poythress et al., 2006).

Skin Measures

The most common skin measure used as an indicator of cognitive workload is based on the moment-to-moment sweat gland and related activities of the autonomic nervous system. When the body perspires, secreted positive and negative ions change the electrical properties of the skin. This phenomenon, originally called the psychogalvanic reflex (PGR), is currently known as the galvanic skin response (GSR), skin conductance response (SCR), or electrodermal response (EDR). In the remainder of this section, it will be referred to as EDR.

EDR can be performed passively or actively. In passive EDR, weak currents generated by the body itself are measured. Active EDR involves applying a small constant current through two electrodes on the skin such that a voltage develops across the electrodes. The skin acts as a variable resistor, and the effective resistance or conductance of the skin can be calculated by applying Ohm’s law (Allanson & Fairclough, 2004). The resulting EDR is expressed in terms of conduction or resistance, which are inversely related (De Waard, 1996).

EDR is frequently conducted actively when measuring cognitive workload. Electrodes are placed where sweat glands are most abundant such as the fingers, palm, forearm, and soles of the feet (Allanson & Fairclough, 2004; De Waard, 1996). In order to determine EDR, a baseline measure, which is typically an average of tonic EDR, is required. Tonic EDR is produced from everyday activities. It varies with psychological arousal and rises when the subject awakens and engages in cognitive effort, especially stress. In contrast, phasic EDR is the result of an external stimulus (De Waard, 1996). One-to-two seconds following an external stimulus, the skin conductance increases and peaks within five seconds (De Waard, 1996; McGraw-Hill Companies, 2007). The amplitude is dependent on the subjective impact of the stimulus, particularly with regard to its novelty and meaning. The wavelike increase in the curve is a fairly reliable indicator of the duration of information-processing and cognitive workload (Collet, Petit, Champely, & Dittmar, 2003).

The advantages of EDR are that it is low-cost and relatively simple to implement. The necessary hardware consists of a low-cost device called a galvanometer and inexpensive electrodes. An amplifier can also be added to increase the signal-to-noise ratio. Often only two electrodes are needed and can be affixed to the skin with adhesive tape.

The main disadvantages to EDR are the latency in response and the insensitivity to cognitive workload when compared to other physiological measures like heart rate variability and EEG. EDR is confounded by the many factors that impact the sympathetic nervous system and the sweat glands, which include temperature, respiration, humidity, age, gender, time of day, season, arousal, and emotions. Because EDR is related to activities of the sympathetic nervous system, all behavior (emotional and physical) can potentially change EDR (De Waard, 1996).

Situation Awareness Techniques

SA_T Calculation

$$SA_T = \frac{\text{Information Acquired}}{\text{Information Required}}$$

For Riese et al.'s work SA_T was based on three elements:

- Location (LOC) - Where is he? Knowing the position of entities or elements to a certain level of accuracy.
- Acquisition (ACQ) - What is he? Knowing the entity or element type to a particular level of acquisition (detect, classify, identify).
- State (STA) - How healthy is he? Knowing the mission-capable status of entities or elements.

The formula they used for the SA_T is below:

$$\text{Instantaneous } SA_T \text{ Score} = \frac{\sum_{i=1}^{n} c_i\left(W_L D_{Li} Loc_i + W_A D_{Ai} Acq_i + W_S Sta_i\right)}{\sum_{i=1}^{n} c_i \bullet (W_L + W_A + W_S)}$$

The SA_T result is between a 0 and 1, where 0 reflects a complete lack of useful SA information and 1 represents full knowledge of location, type, and state of enemy entities within the defined battlespace.

Team SA

Salmon et al (2006) evaluated seventeen SA measures against a set of human factors criteria for use in command, control, communication, computers and intelligence (C4i) environments. They concluded that current SA measurements are inadequate for use in assessing team SA, and recommended a multiple-measure approach. They identified three requirements for C4i SA measurement: ability to simultaneously measure SA at different locations, ability to measure SA in real-time, and ability to measure both individual and team/shared SA (Salmon, Stanton, N. Walker, G., & Green, 2006).

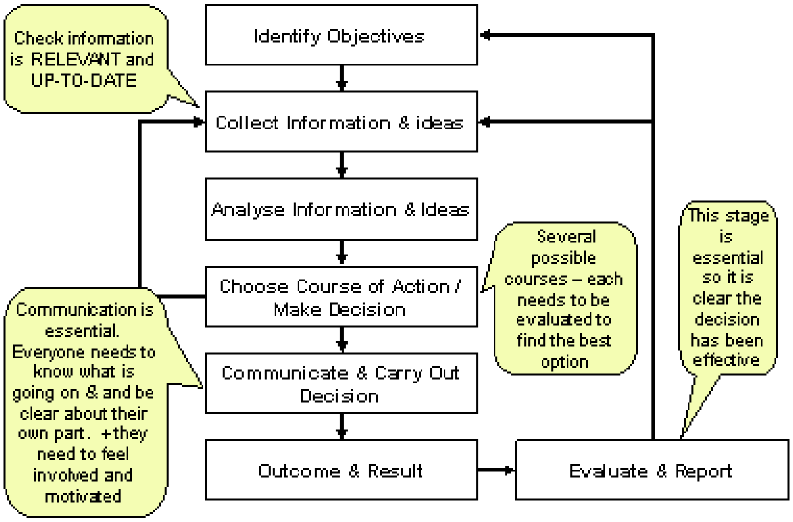
Measurement techniques for team SA have been researched and continue to be a subject of study. Some researchers have aggregated scores from validated SA measurement methods like SAGAT to obtain team SA scores; however, Gorman et al (2006) and others argue that combining individual SA results is not equivalent to team SA because it is missing interaction and critical activities such as coordination, collaboration, and information exchange (Entin & Entin, 2000; Gorman, Cooke, & Winner, 2006). Gorman et al. (2006) developed a measure of SA called the Coordinated Awareness of Situations by Teams (CAST) in which team SA is tested by the group's reaction to unexpected events or "roadblocks." They suggest that these "roadblock" situations are appropriate for determining shared SA because they require the team to coordinate and collaborate in order to circumvent the obstacle. The five steps to CAST involve identifying roadblocks, documenting primary perceptions, documenting secondary perceptions, documenting coordinated perceptions, and documenting coordinated actions (Gorman et al., 2006).

Another team SA technique, Situational Awareness Linked Instances Adapted to Novel Tasks (SALIANT), is similar to CAST. It involves how the team responds to problems, but requires an a priori script and structure. SALIANT includes five phases: 1) identify team SA behaviors (e.g., demonstrated awareness of surrounding environment, recognized problems, anticipated a need for action, demonstrated knowledge of tasks, demonstrated awareness of information) 2) develop scenarios, 3) define acceptable responses, 4) write a script, and 5) create a structured form with columns for scenarios and responses (Gawron, 2000).

Related to team SA is the concept of distributed situation awareness (DSA) (Stanton et al., 2006). Unlike definitions of team SA that examine shared SA, DSA focuses on the system as a whole and consists of the complimentary SA of agents (humans and technology) within the system. Although DSA captures the content of system SA it does not have a technique for assessing the quality of that SA, other than task performance and SME judgment.

Decision Making

In general, the components of decision making, for both system and humans, include identifying objectives, collecting information and ideas, analyzing the information and ideas, choosing a course of action/making a decision, and communicating and carrying out the decision. An outcome results from the action taken and, optimally, there is an evaluation or report reflecting the effectiveness of the decision (Figure 6).

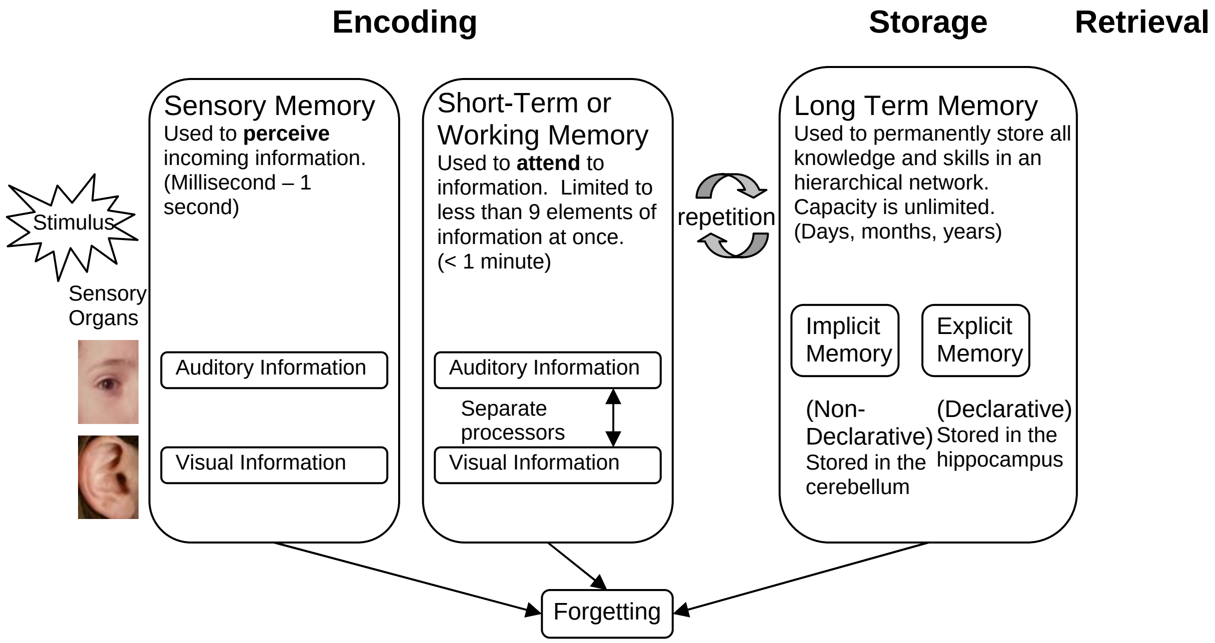

Check information is RELEVANT and UP-TO-DATE
Identify Objectives
Collect Information & ideas
Analyse Information & Ideas
Choose Course of Action / Make Decision
Communicate & Carry Out Decision
Outcome & Result
Evaluate & Report
Several possible courses – each needs to be evaluated to find the best option
This stage is essential so it is clear the decision has been effective
Communication is essential. Everyone needs to know what is going on & and be clear about their own part. +they need to feel involved and motivated

(Finntrack,)

Figure 6: General components of decision making

Decision making depends in part on the system providing the “right” information at the “right” time to the “right” person. In order to achieve this goal, an Internet-like system called the Global Information Grid (GIG) is in the conceptual design phase for C2. The Department of Defense defines the GIG as a “globally interconnected, end-to-end set of information capabilities, associated processing, storing, disseminating, and managing information.” The GIG will be a repository for military information with the goal of providing information superiority over adversaries. A significant risk of the GIG is inundating warfighters with information and providing noisy data that distracts from mission critical data (Bass & Baldwin, 2007). Since net-centric, GIG-type environments are slated for the future, many papers have been written about the GIG and its potential implications. Bass and Baldwin (2007) proposed some rules to direct GIG information flow because information presented at the wrong time, level of detail, and/or lacking proper analysis and interpretation could have devastating effects. Their basic solution is to limit data access to those with authorization and limit data automatically sent to all users (Bass & Baldwin, 2007). The goal is to present a manageable amount of information to warfighters that pertains to them, but also provides context.

Cognitive aspects of decision making begin when the decision maker attends to information in his/her environment. The brain selects data for cognitive processing and encodes data for storage and later retrieval to support decision making (Figure 7). Encoding, translating stimuli to internal (mental) representations, is the longest component of reasoning and decision making taking approximately 45% of the overall time.



Encoding
Storage
Retrieval
Stimulus
Sensory Organs
Sensory Memory
Used to **perceive** incoming information. (Millisecond – 1 second)
Auditory Information
Visual Information
Short-Term or Working Memory
Used to **attend** to information. Limited to less than 9 elements of information at once. (< 1 minute)
Auditory Information
Separate processors
Visual Information
repetition
Long Term Memory
Used to permanently store all knowledge and skills in an hierarchical network. Capacity is unlimited. (Days, months, years)
Implicit Memory
Explicit Memory
(Non-Declarative) Stored in the cerebellum
(Declarative) Stored in the hippocampus
Forgetting

Composed from http://www.scientificjournals.org/journals2007/articles/graphics/1038.gif, http://thebrain.mcgill.ca/flash/i/i_07/i_07_p/i_07_p_tra/i_07_p_tra_2a%20copy.jpg, and http://education.arts.unsw.edu.au/staff/sweller/clt/images/CLT_NET_Aug_97_HTML6.gif

Figure 7: Cognitive Aspects of Decision Making

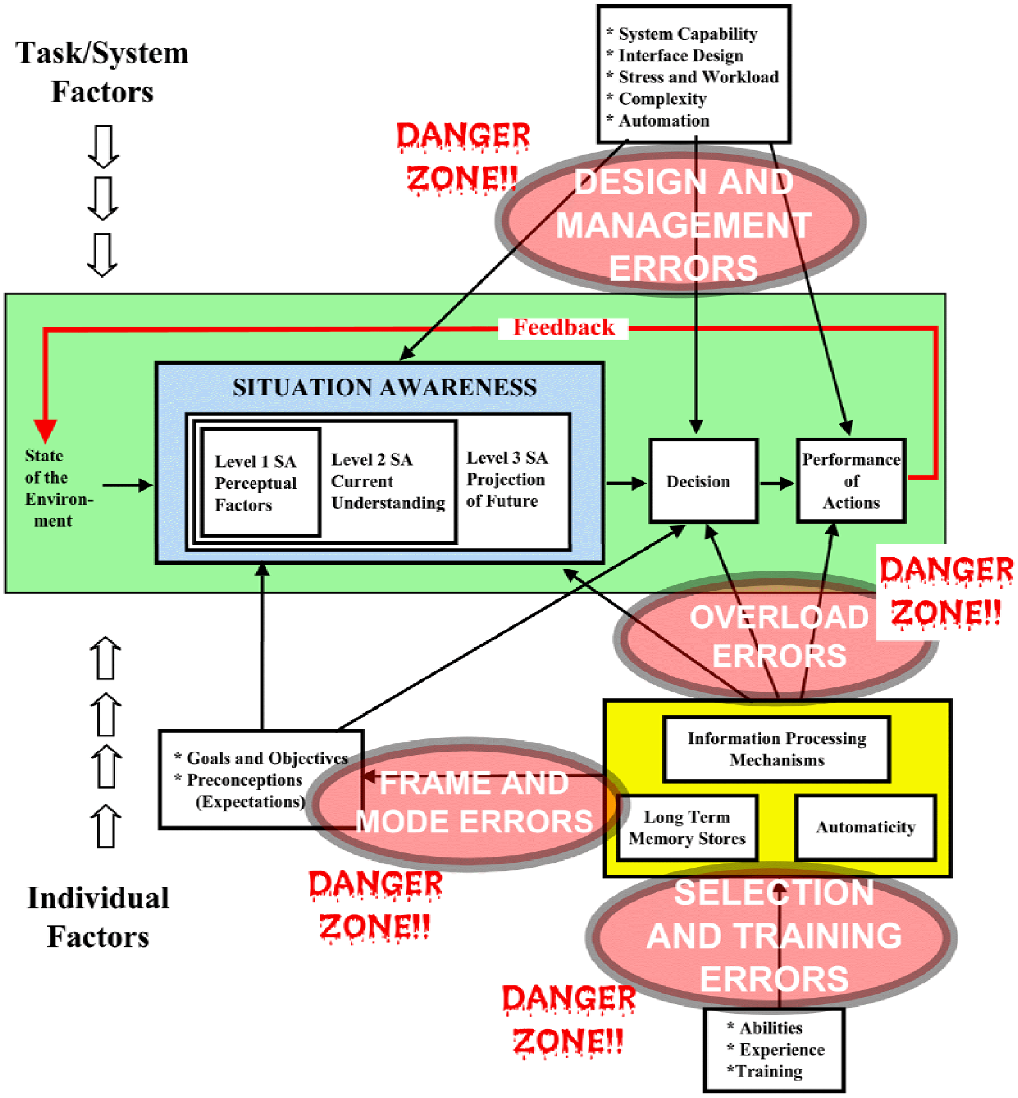
Encoding words takes longer and requires more workload than encoding schematic pictures, so reducing the amount of text will lead to faster decision making. Time-critical decision making displays should aim to decrease encoding time, for example by increasing the intuitiveness of symbology and tasking (Azuma, Daily, & Furmanski, 2006).

The encoding, storage, and retrieval process reflects the close connections between cognitive workload, situation awareness, and decision making. Part of encoding involves perceiving and comprehending, which are critical components of workload and situation awareness. Cognitive workload is related to the storage process because if encoding takes longer or if the information is encoded incorrectly, it may be because the information is more difficult and requires more cognitive resources to process. SA is also an integral part of the storage process because it directly involves perception and comprehension. Typically, cognitive workload and SA are inversely proportionate: a decrease in workload often results in an increase in SA and vice versa. Hence, if the storage process is improved, there is a higher likelihood that cognitive workload will decrease, situation awareness will increase, and decision making will improve as long as the information is relevant to the decision makers' decision.

The projection aspect of SA provides the foundation for decision making. Humans are adept at detecting and remembering patterns. When they notice trends, they begin to extrapolate from the trends to anticipate what will happen next. Series of trends over time result in the development of mental models. This process is called sensemaking. The Command and Control Research Program's Sensemaking Symposium Final Report defines sensemaking as, "the process of creating situation awareness in situations of uncertainty" (Leedom, 2001). Using a data/frame theory as an analogy, data are associated to form a hypothesis called a frame. Preserving and elaborating the frame are like Piaget's assimiliation and reframing is like Piaget's accommodation. Once anchors are established to generate a useful frame, the frame can be evaluated, reframed, elaborated, or compared with alternative frames (G. Klein, Moon, & Hoffman, 2006b). Sensemaking is similar to Endsley's model of situation awareness except instead of a knowledge state that's achieved, sensemaking is the process of getting to the outcome, the strategies, and the obstacles. "Sensemaking is a motivated, continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories, and act effectively" (G. Klein, Moon, & Hoffman, 2006a). Decision makers apply their mental models to the tasks they conduct. Figure 8 provides a model of the relationships between SA and decision making (DM) aspects. In addition, it shows the influence of both the technical/system environment as well as the individual. Finally, it highlights the many areas that can produce errors.

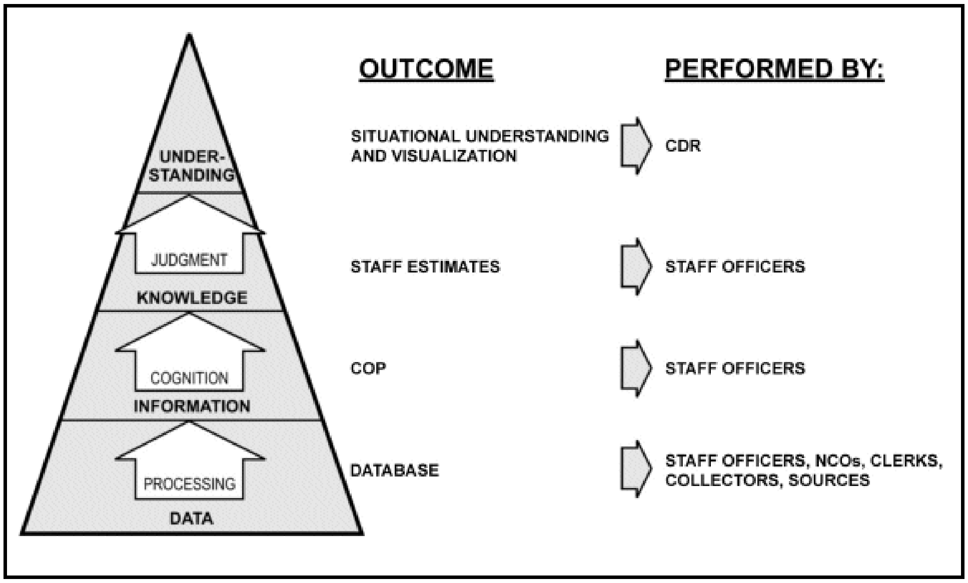
There are several decision making modeling techniques to emulate human decision making. For instance, speeding up decision time provides a tactical advantage in line with John Boyd's high-level sequential model of decision making (consisting of Observation, Orientation, Decision, and Action) called the OODA loop, in which the basic C2 strategy is to execute your OODA loop faster than your opponent (Azuma et al., 2006). Several cognitive models in addition to John Boyd's OODA loop have been developed, for example, ACT-R, EPIC, GOMSL, Soar, IMPRINT, and Bayesian networks (Endsley, Hoffman, Kaber, & Roth, 2007).

Cognitive models tend to generalize decision making; and in uncertain and less familiar situations like complex C2 environments, individual differences, particularly experience, become more apparent (Lamar, 2006). Experience level and prior exposure to similar tasks are discriminating factors between individuals. With novel incoming stimuli, individuals are biased by their past experience, values, and repercussions of the decision (i.e., the cost-benefit analysis) when choosing a course of action (Lamar, 2006). Rasmussen and Reason described experience using a skill-, rule-, and knowledge-based (SRK) classification. Individuals at a skill level have a great deal of experience with the task and the environment and behave virtually automatically. Rule based behavior involves less experience than the skill level and follows an "if, then" process. Knowledge based behavior is unskilled and requires more effort because less memory can be leveraged. Figure 9 below shows an application of SRK based classification to C2.



(Smith, 2007)

Figure 8: Relationship of SA and DM aspects



(Department of the Army, Apr 2003)

Figure 9: SRK Classification of C2

Another individual difference that affects decision making is physiology. Some individuals are innately better and quicker at decision making than others, but aspects of decision making can be learned. For instance, taxi drivers who memorized a map had a different degree of hippocampal volume, a space involved with learning and memory, when compared to those who had not (Lamar, 2006). Expertise leads to differences in function and structure of brain regions required for decision making. Thus, decision making is influenced by both nature and nurture. Another individual difference is personality, but research is ambiguous as to the effect of personality on decision making (Koterba, 2004). Another significant individual difference in decision making is confidence (Sniezek, 1992). Confidence can have critical consequences as disasters frequently result when decision makers are confidently wrong. Conversely, if decision makers have reached an accurate decision and don't have the confidence in their decision to follow-through, the potential benefits would be lost. (Sniezek, 1992)

Most individual difference effects are superseded by trained decision making strategies and short-cuts in time-critical environments like C2 (Koterba, 2004; Lehner, Seyed-Solorforough, O'Connor, Sak, & Mullin, 1997). C2 decision makers often have to follow protocol and procedures they have been trained to execute (Lehner et al., 1997). When C2 decision makers do not have protocol or procedures to follow, they apply rules of thumb or "error-prone heuristics" that result in "good enough" versus exact results obtained from rational decision making outputs (Albers, 1996). Decision makers are especially prone to applying heuristic decision processing under stressful conditions (Lehner et al., 1997). This is a natural tendency because humans have limitations on their capacity to process information, and cope by grouping information and applying mental shortcuts. Kahneman and Tversky pioneered research in this area and began compiling individual cognitive and personal biases in decision making that currently are part of a long list. Some examples relevant to C2 are below:

- Availability Bias: Tendency to overestimate usual or easy to remember events (Lehner et al., 1997)

- Recency Bias: Emphasize recent information; tend to forget or ignore older data (??, 2007)
- Short evidence search: Accept the first alternative that seems credible (??, 2007)
- Source credibility bias: Accept or reject information depending on personal relations with source (??, 2007)
- Ascription of causality: Over generalize correlation as causation (??, 2007)
- Fundamental Attribution Error: View successes as results of talent and failures due to external factors or bad luck, attribute the success of others to good luck and failures to their mistakes (??, 2007; Lehner et al., 1997)
- Hindsight Bias: After an answer is known, people suggest that they knew all along, when they were unclear at the onset (Lehner et al., 1997)
- Choice-supportive bias: Disregard negative aspects of chosen and highlight negative aspects of rejected options to justify choices (??, 2007)
- Inertia: Unwilling to change past thought patterns for new situations (??, 2007)
- Role Fulfillment (Self Fulfilling Prophecy): Causing what is believed will happen to happen (??, 2007)
- Group Think: Willingness to conform to popular group opinion (??, 2007)
- Underestimating uncertainty and the illusion of control: Employ an overly optimistic perception of personal control and not place enough emphasis on the degree of uncertainty (??, 2007)
- Automation Bias: Over reliance on automation (Cummings, Bruni, Mercier, & Mitchell, 2007)
- Anchoring and adjustment: Initial information influences choice (Tatarka, 2002)
- Confirmation bias: find evidence that supports preconceived conclusion and disregarding evidence that contradicts conclusion (Lehner et al., 1997; Tatarka, 2002)

An article from the Military Intelligence Professional Bulletin reports that a large amount of anecdotal evidence suggests the two most dangerous and common biases in military C2-like operations are anchoring and adjustment and confirmation biases (Tatarka, 2002). Once an intelligence analyst has anchored (anchoring and adjustment) on an enemy course of action, they seek evidence that confirms their decision and disregard conflicting information (confirmation bias). The authors suggest that military doctrine promotes this bias because of short time constraints. The risk is “cognitive tunnel vision”, which is emphasized in high stress situations like C2, and could lead to devastating effects (Tatarka, 2002). Another potentially dangerous and common heuristic in C2 is automation bias. As technology becomes more prevalent and can provide automated solutions, operators may over rely on the automation, become complacent, and may experience a loss of situation awareness (Cummings et al., 2007). These cognitive biases are highly relevant to consider in decision making tasks because they become increasingly resistant to preventative techniques such as, training, decision support tools, and devil’s advocate approaches, in unfamiliar and stressful environments like complex C2 (Lehner et al., 1997; Tatarka, 2002). The heuristics and biases can be better understood by applying techniques like those in understanding information requirements; for example, cognitive task analysis, cognitive work analysis, etc.

CTA Methodologies

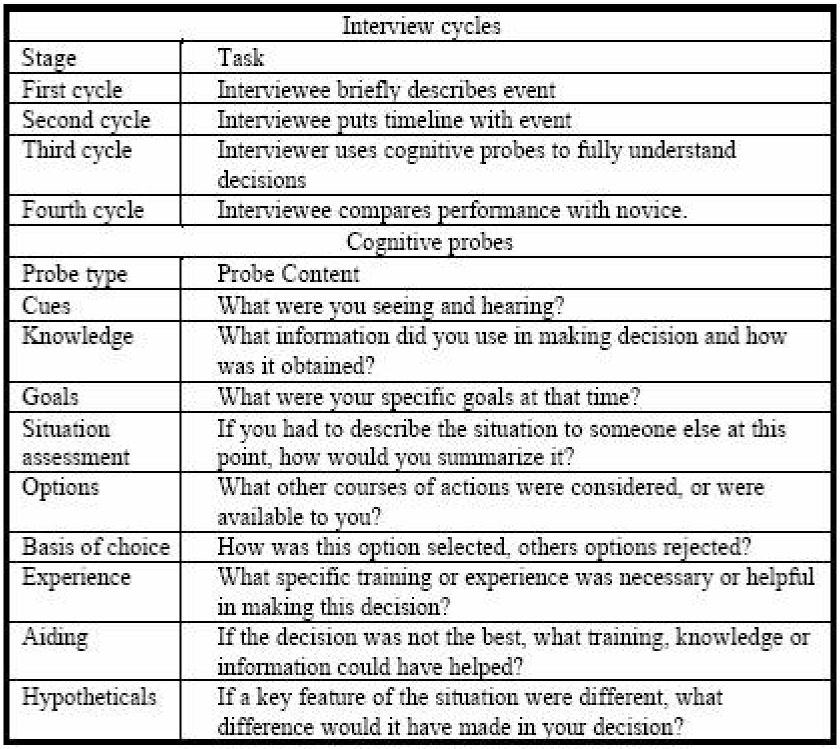
A number of methods exist to obtain information requirements and decision rationale from the decision maker, the most common being cognitive task analysis (CTA). A variety of CTA methodologies have been developed that differ in approach, structure, emphasis, and resource requirements, but all include some sort of knowledge elicitation, analysis, and knowledge representation (Federal Aviation Administration Human Factors Division, 1999; Militello & Hutton, 1998). Often the knowledge elicited is not measured quantitatively; however, aspects that can be measured quantitatively are number or type of ideas considered and the number or type of consequences considered. The analysis and knowledge representation are the significant parts of the CTA as they can be used to improve processes or systems. If the system and process are in line with the decision makers' mental model derived from conducting a CTA, there is a higher likelihood that the decision maker will be able to use the system more effectively and efficiently to make a decision.

Three examples of CTA are: 1) Precursor, Action, Results, and Interpretation (PARI) method 2) Critical Decision Method (CDM), and 3) Conceptual Graph Analysis (CGA). In the PARI method, subject matter experts identify troubleshooting use cases to elicit system knowledge (how the system works), procedural knowledge (how to perform problem solving procedures), and strategic knowledge (knowing what to do and when to do it) from other subject matter experts (Federal Aviation Administration Human Factors Division, 1999; Jonassen, Tessmer, & Hannum, 1998). "PARI attempts to identify each *A*ction (or decision) that the problem solver performs, the *P*recursor (or Prerequisite) to that action, the *R*esult of that action, and an expert's *I*nterpretation of the Results of that Action" (Jonassen et al., 1998). The PARI technique, developed by Hall, Gott, and Pokorny in 1995, was originally designed for the Air Force to design intelligent tutoring systems. It involves a structured interview during which pairs of subject matter experts probe each other under realistic conditions. The interviews are held during and after troubleshooting has occurred with an emphasis on reasoning used in making decisions. PARI products include flowcharts, annotated equipment schematics, and tree structures (Federal Aviation Administration Human Factors Division, 1999).

An advantage of PARI is that it thoroughly exposes how subject matter experts deal with systems by identifying the technicalities of how the system works (technology focused), how to perform problem solving procedures (human-system interface), and knowing what to do about the problem (cognitive or decision making rationale). PARI is especially strong in revealing troubleshooting and analyzing problem solving techniques that can be beneficial for training. The disadvantage of PARI is that it may focus too much on specific troubleshooting and it relies heavily on subject matter expertise. Since PARI consists of subject matter experts interviewing each other, there is a risk that some details may be left out because they may be assumed to be included or considered trivial.

CDM, based on Flanagan's critical incident technique developed in 1954, and formalized by Klein in 1993, is a series of semi-structured interviews of subject matter experts that focuses on critical, non-routine incidents requiring skilled judgment (G. A. Klein, Calderwood, & Macgregor, 1989). The interview is considered semi-structured because it is in between an ongoing verbal protocol where the decision maker "thinks aloud" and a completely structured interview (G. A. Klein et al., 1989). The theory behind CDM is that probing subject matter experts about difficult activities results in the "richest source of data" to understand decision making of highly skilled personnel as the information gleaned is expertise, not formalized protocol (G. A. Klein et al., 1989). When CDM is conducted, subject matter experts recount a difficult incident and the interviewer probes to distinguish decision points, critical cues, cognitive strategies, etc. (Table 1 provides information on interview cycles and example cognitive probes.) (Federal Aviation Administration Human Factors Division, 1999) p.126.

Table 1: CDM Interview Cycles and Cognitive Probes



Interview cycles	
Stage	Task
First cycle	Interviewee briefly describes event
Second cycle	Interviewee puts timeline with event
Third cycle	Interviewer uses cognitive probes to fully understand decisions
Fourth cycle	Interviewee compares performance with novice.
Cognitive probes	
Probe type	Probe Content
Cues	What were you seeing and hearing?
Knowledge	What information did you use in making decision and how was it obtained?
Goals	What were your specific goals at that time?
Situation assessment	If you had to describe the situation to someone else at this point, how would you summarize it?
Options	What other courses of actions were considered, or were available to you?
Basis of choice	How was this option selected, others options rejected?
Experience	What specific training or experience was necessary or helpful in making this decision?
Aiding	If the decision was not the best, what training, knowledge or information could have helped?
Hypotheticals	If a key feature of the situation were different, what difference would it have made in your decision?

A variety of CDM products can be produced; one of the most common is a narrative account (Federal Aviation Administration Human Factors Division, 1999). Another product is a cognitive requirements table that includes cognitive demands of the task and pertinent contextual information. CDM results are usually used to develop system design recommendations or training (Federal Aviation Administration Human Factors Division, 1999).

CDM was implemented in a C2 decision making study of anti-air warfare operators on a U.S. Navy AEGIS cruiser to investigate decision maker strategies. Results reflected that the feature matching strategy, involving recognition of a typical class of situation, was the most used strategy (87% of diagnostic strategies). Story building was also used (12% of diagnostic strategies) where the situation was novel or where the decision maker builds a story from seemingly disparate pieces of information to develop a coherent explanation of the situation. Decision makers did not evaluate 75% of the decisions that they implemented, and considered and compared multiple options in only 4% of the cases. In the 4% of cases where multiple options were considered, they were not the most critical decision points. When decision makers did not understand a situation, they prepared for the worst case scenario probably to avoid risk (Kaempf, Klein, Thordsen, & Wolf, 1996).

The advantage of CDM is that it reveals expertise and understanding of objectives that would not otherwise be illuminated. The semi-structured organization provides flexibility to the decision maker to discuss aspects that might not have been specified a priori. It has also been used to in complex C2 environments to determine decision making strategies. The interview cycle approach expands the attributes of information collected, and also increases the time and resources required to conduct CDM. A disadvantage of CDM is that it is subjective and reflective on the decision maker's own strategies and basis for decisions (G. A. Klein et al., 1989). Another disadvantage is that the critical event chosen may be very atypical or rare. Finally, since CDM is less structured, it is more difficult to interpret and analyze the results.

CGA was developed in 1992 by Gordon and Gill and involves generating a visualization of conceptual knowledge structures to conduct CTA (Federal Aviation Administration Human Factors Division, 1999). A CGA consists of a formal and detailed collection of nodes (which can be goals, actions or events), relations, and questions (Federal Aviation Administration Human Factors Division, 1999; Jonassen et al., 1998). Nodes are connected via arcs, which portray the relationship between nodes. The CGA process begins by exploring any pre-existing documentation related to the task to be analyzed. Then, a process called free generation is implemented in which an SME leverages the existing documentation and adds task information requirements. The information is then compiled and visually presented as a draft conceptual graph. Any gaps in the representation are constructed into detailed questions. If there are still gaps after questions are asked, information is filled in from observations (Federal Aviation Administration Human Factors Division, 1999). The last step is validating the conceptual graph by having an expert perform the task and check for incorrect or missing information (Jonassen et al., 1998).

An advantage to CGA is that it provides a visual depiction of internal knowledge like a concept map. Clarifying the linkages between concepts causes the interviewer to closely investigate the conceptual relationships that might not be examined through other CTA techniques. Another advantage is the detailed approach affords a systematic process with more structure than other CTA methods. The structure also yields "specific yet comprehensive" questions. In addition, a variety of automated software tools exist to assist in developing conceptual graphs like COG-C. A disadvantage is the CGA nodes and arcs take time to learn, and a CGA is difficult to develop while an unstructured interview is taking place. Also, while CGA describes concepts well, it is weak at capturing procedural knowledge (Jonassen et al., 1998).

Despite the contrast between CTA techniques, they are useful in revealing C2 decision maker rationale. The PARI technique is considered a traditional cognitive task analysis technique, CDM is considered activity-based analysis, and CGA is considered subject matter/content analysis; however, all reveal information that could improve C2 processes, or be used to evaluate C2 decision making. Most of these techniques focus on deviations from standard operating procedures and preplanned responses where C2 decision making and expertise can be exposed (Jonassen et al., 1998). In addition, the various levels of structure in CTA methodologies parallels the levels of structure in various aspects of C2 (G. A. Klein et al., 1989). Another positive aspect is that many of the CTA techniques are conducted retrospectively, which is important in C2 because they are less intrusive (G. A. Klein et al., 1989). The caveat to CTAs are that "no well-established metrics exist" for evaluating CTAs, and it is difficult to evaluate differences between CTA methods (Militello & Hutton, 1998). This is partially because it is unknown what information is lost versus gained in comparison to other techniques and also because interviewees and individuals provide different information each. Also, CTAs can be very resource intensive. Because individual differences impact how much information individuals are willing to provide and respond, it is difficult to assess the reliability and validity of CTA methods. Also, another caveat is that no advanced techniques for team CTA have been developed (Militello & Hutton, 1998).

References

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics, 36,* 623-636.

Albers, M. J. (1996). Decision making: A missing facet of effectivfe documentation. *ACM Special Interest Group for Design of Communication: Proceedings of the 14th Annual International Conference on Systems Documentation: Marshaling New Technological Forces: Building a Corporate, Academic, and User-Oriented Triangle,* , 57-65.

Allanson, J., & Fairclough, S. H. (2004). A research agenda for physiological computing. *Interacting with Computers, 16*(5), 857-878.

Azuma, R., Daily, M., & Furmanski, C. (2006). A review of time critical decision making models and human cognitive processes. *Aerospace Conference, 2006 Institute of Electrical and Electronics Engineers, Inc.,*

Basar-Eroglu, C., & Demiralp, T. (Jan 2001). Event-related theta oscillations: An integrative and comparative approach in the human and animal brain. *International Journal of Psychophysiology, 39*(2-3), 167-195.

Bass, S. D., & Baldwin, R. O. (2007). A model for managing decision-making information in the GIG-enabled battlespace. *Air and Space Power Journal,* , 100-108.

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovis, M. M., Davis, G., Lumicao, M. N., et al. (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.

Berka, C., Levendowski, D. J., Ramsey, C. K., Davis, G., Lumicao, M. N., Stanney, K., et al. (2005). Evaluation of an EEg-workload model in the aegis simulation environment. Paper presented at the *, 5797* 90-99.

Boehm-Davis, D. A., Gray, W. D., & Schoelles, M. J. (2000). The eye blink as a physiological indicator of cognitive workload. *Proceedings of the IEA 2000/HFES 2000 Conference,*

Collet, C., Petit, C., Champely, S., & Dittmar, A. (2003). Assessing workload through physiological measurements in bus drivers using and automated system during docking. *Human Factors, 45*(4), 539-548.

Colman, A. M. (2001). A dictionary of psychology: P300. ()Oxford University Press.

Cummings, M. L., Bruni, S., Mercier, S., & Mitchell, P. J. (2007). Automation architecture for single operator-multiple UAV command and control. *The International C2 Journal: Special Issue- Decision Support for Network-Centric Command and Control, 1*(2), 1-24.

Dantest Medical System. *What is heart rate variability (HRV) analysis?* Retrieved 12/4, 2008, from http://www.dantest.com/introduction_what_is_hrv.htm

De Waard, D. (1996). The measurement of drivers' mental workload. (PhD thesis, University of Groningen).

Department of Kinesiology University of Waterloo. *Parietal lobe.* Retrieved 1/13, 2008, from http://ahsmail.uwaterloo.ca/kin356/dorsal/parietal.jpg

Department of the Army. (Apr 2003). *Field manual 3-21.21: The stryker brigade combat team infantry battalion*

Endsley, M. R., Hoffman, R., Kaber, D., & Roth, E. (2007). Cognitive engineering and decision making: An overview and future course. *Journal of Cognitive Engineering and Decision Making, 1*(1), 1-21.

Entin, E. B., & Entin, E. E. (2000). Assessing team situation awareness in simulated military missions. (1) 73-76.

Federal Aviation Administration Human Factors Division. (1999). *Department of defense handbook: Human engineering program process and procedures* No. MIL-HDBK-46855A)

Finntrack. *Decision making process.* Retrieved 2/12, 2008, from http://www.finntrack.com/hnc_hnd/bus-decision.htm

Gawron, V. J. (2000). *Human performance measures handbook*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science, 4*(1-2), 113-131.

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics, 49*(12-13), 1312-1325.

Hannula, M., Koskelo, J., Huttenen, K., Sorri, M., & Leino, T. (2007). Artificial neural network analysis of heart rate under cognitive load in a flight simulator. Paper presented at the

Harris, B. *A revolution in neuroscience: Tuning the brain.* Retrieved 1/13, 2008, from http://www.centerpointe.com/about/articles_research.php

Jonassen, D. H., Tessmer, M., & Hannum, W. H. (1998). *Task analysis methods for instructional design*

Kaempf, G. L., Klein, G., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors, 38*(2), 220-231.

Klein, G., Moon, B., & Hoffman, R. R. (2006a). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems, 21*(4), 70-73.

Klein, G., Moon, B., & Hoffman, R. R. (2006b). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems, 21*(5), 88-92.

Klein, G. A., Calderwood, R., & Macgregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics, 19*(3), 462-472.

Koterba, N. T. (2004). *APL internal report: The effect of personality on decision making*

Lamar, M. (2006). Neuroscience and decision making.

Leedom, D. K. (2001). *Sensemaking symposium final report*Command and Control Research Program.

Lehner, P., Seyed-Solorforough, M., O'Connor, M. F., Sak, S., & Mullin, T. (1997). Cognitive biases and time stress in decision making. *IEEE Transactions on Systems, Man, and Cybernetics- Part A: Systems and Humans, 27*(5), 698-703.

McGraw-Hill Companies, I. (2007). *Science and technology encyclopedia, 5th edition: Electrodermal response.* Retrieved 5/18, 2007, from http://www.answers.com/topic/galvanic-skin-response

Militello, L. G., & Hutton, R. J. B. (1998). Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 41*(11), 1618-1641.

Nickel, P., & Nachreiner, F. (2003). Sensitivity and diagnosticity of the 0.1-hz component of heart rate variability as an indicator of mental workload. *Human Factors, 45*(4), 575-590.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. Boff, L. Kaufman & J. Thomas (Eds.), *Handbook of perception and performance* (vol. 2 ed.,). New York: Wiley.

Poole, A. (2004). Eye movements.

Poythress, M., Russell, C., Siegel, S., Tremoulet, P. D., Craven, P., Berka, C., et al. (2006). Correlation between expected workload and EEG indices of cognitive workload and task engagement. *2nd Annual AugCog International Conference*, San Fransisco, CA. 32-44.

Rowe, D. W., Silbert, J., & Irwin, D. (1998). Heart rate variability: Indicator of user state as an aid to human-computer interaction. Paper presented at the 480-487.

Salmon, P., Stanton, N. Walker, G., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics, 37*, 225-238.

Smith, D. J. (2007). *Situation(al) awareness in effective command and control.* Retrieved 1/3, 2008, from http://www.smithsrisca.demon.co.uk/situational-awareness.html

Sniezek, J. (1992). Groups under uncertainty: An examination of confidence in group decision making. *Organizational Behavior and Human Decision Making Processes,* (52), 124-155.

Tatarka, C. J. (2002, Overcoming biases in military problem analysis and decision-making. [Electronic version]. *Military Intelligence Professional Bulletin,* (Jan-March, 2002)

Tattersall, A. J., & Hocky, G. R. J. (1995). Level of operator control and changes in heart rate variability during simulated flight maintenance. *Human Factors, 37*(4), 682-698.

Tsai, Y., Viirre, E., Strychacz, C., Chase, B., & Jung, T. P. (2007). Task performance and eye activity: Predicting behavior relating to cognitive workload. *Aviation, Space, and Environmental Medicine: Operational Applications of Cognitive Performance Enhancement Technologies,*

Van Orden, K. F. (2000). *Real-time workload assessment and management strategies for command and control watchstations: Preliminary findings.* Unpublished manuscript. Retrieved March 16, 2006, from http://www.dtic.mil/matris/sbir/sbir011/Navy89b.doc

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors, 1*, 111-121.

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics, 41*(5), 656-669.

Wickens, C. D., Mavor, A. S., & McGee, J. P. (Eds.). (1997). *Flight to the future: Human factors in air traffic control*National Academy Press.

APPENDIX

Workload Measurement Techniques

Subjective

NASA TLX, the most commonly used subjective workload scale, contains six dimensions: 1) Mental demand, which refers to perceptual and cognitive activity, 2) Physical demand, which refers to physical activity, 3) Temporal demand, which refers to time pressure, 4) Performance, which is related to personal goal accomplishment, 5) Effort, which refers to energy expenditure in accomplishing the required level of performance, and 6) Frustration, which is related to feelings of irritation, stress, etc. Subjects complete a pair-wise comparison procedure to weight the dimensions before the task. After completing the task, subjects rate the six dimensions using a 0-100 scale (Hart & Staveland, 1988).

The SWAT contains three dimensions: 1) Time (T) load, which reflects the amount of spare time available in planning, executing, monitoring a task; 2) Mental effort (E) load, which assesses how much conscious mental effort and planning are required to perform a task; and 3) Psychological stress (S) load, which measures the amounts of risk, confusion, frustration, and anxiety associated with task performance. The three SWAT dimensions (i.e. T, E, and S) are at three discrete levels (i.e. 1, 2, and 3). Subjects rank the three dimensions and three levels by perceived importance in a 3^3 or 27-card sorting exercise before conducting the task. The card sort results in seven weighting schemes: TES, ETS, SET, TSE, EST, STE, and equal emphasis on T, E, and S. After completing the task, subjects rate the task on the three dimensions (Reid & Nygren, 1988).

Wierwille and Casali (1983) modified the wording of the validated physically focused Cooper-Harper Rating Scale such that it would be appropriate for assessing cognitive functions like “perception, monitoring, evaluation, communications, and problem solving.” The MCH scale maintains a decision tree architecture where participants respond to yes or no questions that lead to options for rating. The rating scale is 1-10, where 1 is very easy and 10 is impossible (O'Donnell & Eggemeier, 1986).

Physiological

Eye Measures

A variety of studies have shown that various aspects of eye behavior correlate with cognitive workload. One of the most sensitive eye physiological measures is pupil diameter. Pupil diameter increases (dilates) as cognitive workload increases (Ahlstrom & Friedman-Berg, 2006). Pupil diameter changes can be dynamic, for instance during comprehension of individual sentences, or sustained during recall of digit span (Van Orden, 2000). Although the average pupil diameter changes by as much as 0.6mm when recalling seven digits, many confounds such as ambient lighting, stimulus characteristics, and even emotional effects can cause pupillary responses that are greater than those from workload alone (O'Donnell & Eggemeier, 1986; Van Orden, 2000). Also, accurate measurement techniques required may impose constraints on experimentation by requiring the subject to stay in one location or wear a measuring device on his/her head. Pupil diameter measurements are therefore difficult to use in applied settings where the environment and other external factors are not controlled. Finally, research suggests that pupil diameter measurements may be highly responsive to cognitive workload changes, yet not diagnostic because there is little ability to identify the resource (e.g. visual, auditory, etc.) utilized in the task (O'Donnell & Eggemeier, 1986).

Generally, blink rate and blink duration decrease as workload increases (Ahlstrom & Friedman-Berg, 2006; Van Orden, 2000; Veltman & Gaillard, 1998). Boehm-Davis et al. (2000) suggest that eye blinks are suppressed when individuals are engaged in cognitive processing; however, eye blinks show great variability (Boehm-Davis, Gray, & Schoelles, 2000; O'Donnell & Eggemeier, 1986). Blink duration is also somewhat unreliable to gauge cognitive workload because other factors like visual workload confound cognitive workload. A visual tracking task with minimal cognitive load can cause lower blink durations than during a more cognitively challenging flight simulation task (Van Orden, 2000). Therefore, eye blinks and blink duration should be considered global indicators of long-term effects versus specified diagnostic techniques (O'Donnell & Eggemeier, 1986).

Another potential eye related measurement of workload is number and duration of saccades. The number of saccades, which are a series of small, quick, jerky movements of the eyes when changing focus from one point to another in the visual field, increase as workload increases. Saccade duration, which typically lasts for 20 to 35 milliseconds, decreases as workload increases. (De Waard, 1996; Poole, 2004; Wickens, Mavor, & McGee, 1997) While saccades may provide clues about the cognitive strategy employed, studies reflect that prior to voluntary eye movement, attention shifts to the location of interest; therefore saccades may be measures of attention versus cognitive workload (Tsai, Viirre, Strychacz, Chase, & Jung, 2007).

After each saccade, the eyes stay still and encode information in movements called “fixations.” Fixation frequency and fixation duration or dwell time both increase as cognitive workload increases, but is task dependent (De Waard, 1996; Poole, 2004; Van Orden, 2000). For example, during a challenging flight simulation, fixation duration correlated with the number of flight rule errors, reflecting a correlation with cognitive workload; however, in a challenging visual search task, search fixation frequency increased and fixation duration did not change (Van Orden, 2000). Fixation is particularly sensitive to visual workload, making it more diagnostic than other techniques; however, fixation does not necessarily imply cognition (De Waard, 1996).

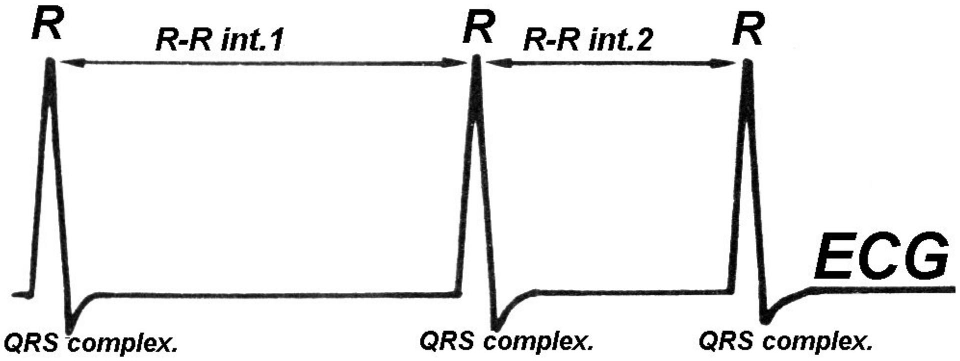
Scan paths, recurring patterns of saccades and fixations, become less of a pattern between display elements as workload increases (O'Donnell & Eggemeier, 1986; Poole, 2004). Also, dwell times in each position lengthen and fewer display elements are used. Scanning is typically a global indicator of workload, but scanning may be a diagnostic index of the source of workload within a multi-element display environment if (1) critical information must be gathered from multiple locations, (2) relative importance of data obtained from each location is different, and (3) the subject can adjust or change the imposed load by a change in strategy (O'Donnell & Eggemeier, 1986).

In summary, blink rate, blink duration, and saccade duration all decrease while pupil diameter, the number of saccades, and the frequency of long fixations all increase with increased workload.

The eye is readily accessible to observation and provides rich data that can be assessed, but study results differ in eye physiological measures that correlate highest with cognitive workload. For example, in a mock anti-air warfare task, blink frequency, fixation frequency, and pupil diameter were the most predictive variables correlating eye activity to target density (Van Orden, 2000); and in an air traffic controller study, managing traffic during adverse weather conditions, decreased saccade distance, blink duration, and pupil diameter correlated closest to cognitive workload (Ahlstrom & Friedman-Berg, 2006). Because of the variability amongst eye physiological measures, it is recommended that multiple techniques be combined (Van Orden, Limbert, Makeig, & Jung, 2001). Several confounds including visual workload impact results, so eye measures should be used as global indicators of cognitive workload. Eye physiological measures provide more sensitive results in controlled environments.

Cardiac Measures

The main cardiac measures studied for sensitivity to cognitive workload include the electrocardiogram, blood pressure, and blood volume. Of these three, measures of electrocardiographic activity show the most promise (Rowe, Silbert, & Irwin, 1998). The electrocardiograph produces a graphic called an electrocardiogram, abbreviated ECG or EKG, from the German *elektrokardiogramm*, which records the electrical activity of the heart over time. Surface electrodes are placed on the skin of a subject to identify pulse beats, recognizable by a pattern called the QRS complex (Figure 1) (O'Donnell & Eggemeier, 1986).



(Dantest Medical System,)

Figure 1: Heart Rate and Heart Rate Variability

Although absolute heart rate has been used as a measure of overall workload, spectral analysis of heart rate variability (HRV) or sinus arrhythmia reflects some correlation to cognitive workload (De Waard, 1996; O'Donnell & Eggemeier, 1986; Tattersall & Hocky, 1995; Veltman & Gaillard, 1998). As the name implies, HRV is the variability in heart rate or the variability between R-R intervals (see Figure 1). Of the over 30 techniques available for determining HRV (e.g. Fourier transform, autoregressive modeling, time-varying analysis, broadband spectral analysis, etc.), most cognitive loading HRV measures emphasize frequency. Frequency HRV techniques measure the amount of variation in different frequency bands. There are three major frequency bands: (1) very low-frequency band (0.0033-0.04 Hz), associated with temperature regulation and physical activity, (2) low-frequency band (0.04-0.15 Hz), associated with short-term regulation of arterial pressure; and (3) high-frequency band (0.15-0.40 Hz), reflecting the influence of respiration. Several studies have suggested that the low-frequency band, and specifically what is called the 0.10 Hz component, indicates cognitive workload (De Waard, 1996; Nickel & Nachreiner, 2003; O'Donnell & Eggemeier, 1986). The 0.10 Hz component reflects short-term changes in blood pressure. A peak of the 0.10 Hz component reflects decreased cognitive workload, and a flattening of the 0.10 Hz component reflects conditions of greater mental workload (Rowe et al., 1998).

Nickel and Nachreiner (2003) assessed the diagnosticity (i.e., the ability to differentiate amongst different types of tasks) and sensitivity (i.e., the ability to detect levels of difficulty) of the 0.1 Hz component of heart rate variability (HRV) for cognitive workload using 14 cognitive tasks (e.g. reaction time, mathematical processing, memory-search, grammatical reasoning task, etc.) from an environmental stressors standardized test in a laboratory context. Only one type of task could be discriminated as different from the other types of tasks – that task reflected a cognitive loading score that matched the cognitive loading expected at rest; however, these results directly conflicted with performance (i.e., performance errors were made) and perceived difficulty (i.e., participants reported mental workload). In terms of sensitivity, the results echoed several other studies that HRV can discern between work and rest, but not to gradations in between (Rowe et al., 1998). Because the experimenters noted differences in the 0.10 Hz component when time pressure was involved, they propose that HRV be used as an indicator for emotional strain or time pressure versus cognitive workload (Nickel & Nachreiner, 2003).

Research by Hannula et al. (2007) supports the use of HRV as a stress indicator. They applied an artificial neural network analysis to evaluate the relationship between cognitive workload that raises the psychophysiological stress and HRV data in fighter pilots. The Pearson's coefficients between the ECG data and the cognitive workload that increases psychophysiological stress as evaluated by an experienced flight instructor were between 0.66 and 0.69 (Hannula, Koskelo, Huttenen, Sorri, & Leino, 2007).

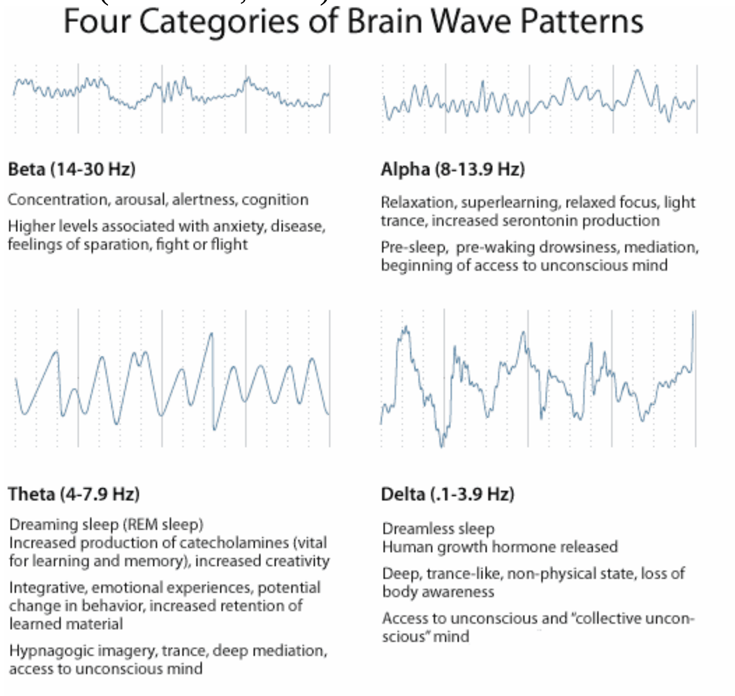
There are several caveats to using heart rate variability as a cognitive workload measure. Possibly the most significant disadvantage to HRV is that it has not been validated as a sensitive cognitive workload indicator (O'Donnell & Eggemeier, 1986). Also, heart rate and, likewise to a smaller extent, HRV are confounded by psychological processes like high responsibility or fear, physical effort and speech, and environmental factors such as high G-forces (De Waard, 1996; O'Donnell & Eggemeier, 1986; Tattersall & Hocky, 1995). As indicated in the studies discussed above, HRV is influenced by stress and time constraints. Physical effort will impact HRV results unless it is kept to a minimum and constant across conditions (De Waard, 1996). Speech can also confound HRV results if verbalization is longer than 10 s and relatively frequent (more than one to five times per minute) (De Waard, 1996). Another factor that affects HR measures, and to a lesser degree cognitive workload, is age. If HRV is the primary workload measure, it may be necessary to restrict elderly subjects from participation because HRV may decrease with increasing age. Finally, a last caveat to consider is that operators typically need to act as their own control because of the idiosyncrasies in the measure (De Waard, 1996).

Despite the caveats, heart rate measurement is arguably the simplest physiological index to measure and it has been employed extensively. The ECG signal requires minimal amplifying (approximately 10 to 20 times less than continuous EEG) and if measurements are limited to R-wave detection and registration, then electrode placement is not critical (De Waard, 1996). Cardiac techniques are also the most popular physiological technique in the last 40 years (Rowe et al., 1998). They are relatively noninvasive and unobtrusive (O'Donnell & Eggemeier, 1986). Also, with continuously recorded cardiac measures, research has shown that HRV can indicate within seconds the change from work to rest (Rowe et al., 1998).

Neural Measures

Electroencephalography, the measurement of electrical activity in the brain, is the most common neurophysiological technique used as an indicator of cognitive workload (Berka et al., 2004). It typically involves a noninvasive procedure of placing electrodes on the surface of the head to detect activity through the skull and scalp. In rare instances, electrodes are placed subdurally or in the cerebral cortex. The traces of activity that result are called an electroencephalogram (EEG), which represents the electrical signals or postsynaptic potentials from a large number of neurons. In clinical use the EEG is considered a “gross correlate of brain activity” because instead of measuring electrical currents from individual neurons, it reflects relative voltage differences amongst various brain areas.

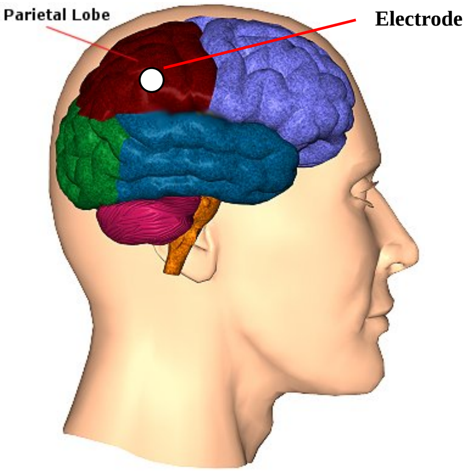
Frequency analyses performed on EEG signals are also called epoch analyses or background EEG analyses and usually result in four ranges or wave patterns (Figure 3) (De Waard, 1996). Although these wave patterns or bands are traditionally used to provide information about the health and function of the brain, some are very responsive to variations in alertness and attention. Specifically, several studies confirm that alpha and especially theta bands are sensitive to aspects of short-term or working memory (Gevins & Smith, 2003). When working memory is in use, research reflects decreases in the upper alpha band and increases in the theta band that become more exaggerated when load increases. This suggests that these bands may be indicators of cognitive loading. One study reflected decreased alpha and increased theta activity during dual-task performance when compared to single-task performance. However, individual differences may be significant; for example, a small number of individuals do not generate alpha waves at all (De Waard, 1996).


Four Categories of Brain Wave Patterns
Beta (14-30 Hz)
Concentration, arousal, alertness, cognition
Higher levels associated with anxiety, disease, feelings of sparation, fight or flight
Alpha (8-13.9 Hz)
Relaxation, superlearning, relaxed focus, light trance, increased serontonin production
Pre-sleep, pre-waking drowsiness, mediation, beginning of access to unconscious mind
Theta (4-7.9 Hz)
Dreaming sleep (REM sleep)
Increased production of catecholamines (vital for learning and memory), increased creativity
Integrative, emotional experiences, potential change in behavior, increased retention of learned material
Hypnagogic imagery, trance, deep mediation, access to unconscious mind
Delta (.1-3.9 Hz)
Dreamless sleep
Human growth hormone released
Deep, trance-like, non-physical state, loss of body awareness
Access to unconscious and "collective unconscious" mind

(Harris,)

Figure 3: Four Categories of Brain Wave Patterns

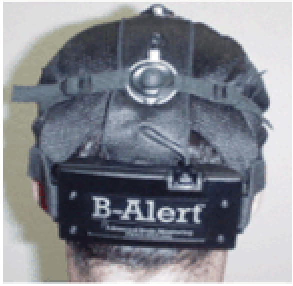
In addition to background EEG, event-related potentials (ERPs) are a neural measurement technique used as a cognitive workload indicator (Berka et al., 2004). Unlike evoked potentials, which are the result of physical stimuli, ERPs may be caused by processes such as memory, attention, expectation, or changes in mental state(Basar-Eroglu & Demiralp, Jan 2001). ERPs are any response to an internal or external stimulus, but are often considered to be the direct result of cognitive activity. ERPs can be reliably measured from voltage deflections in EEG. For instance approximately 300 milliseconds following unpredictable stimuli, there are positive deflections in voltage called the P300 or more simply, P3. The P3 peaks around 300 ms for very simple decisions, and the amplitude increases with unexpected, task-relevant stimuli. After 300 ms, the latency of P3 increases with task difficulty and relates to cognitive workload (De Waard, 1996). During primary tasks only, the P3 amplitude increases with task complexity; however when secondary tasks are added, the P3 decreases with primary task complexity (De Waard, 1996). An external factor is age because the P3 latency increases between 1 and 2 milliseconds each year of the adult lifespan (Colman, 2001). P3s are somewhat difficult to detect, and therefore it is often necessary to present an individual stimulus dozens or hundreds of times and average the results together to cancel out the noise and present the stimulus response clearly. The signal to noise ratio can be improved by placing electrodes on the participant's head above the parietal lobe of the brain (Figure 4).



(Department of Kinesiology University of Waterloo,)

Figure 4: Parietal sites for electrodes

The B-Alert System (Figure 5), a device developed by Advanced Brain Monitoring (ABM), is a commercially available product that claims to have a cognitive workload measurement capability. Originally created to provide early detection of drowsiness, B-Alert advertises real-time workload calculations via a wireless, light weight EEG headset that can analyze six channels of EEG. The signal processing including amplification, digitization, and radio frequency transmission of the signals, is built into a portable unit worn with the headset (Berka et al., 2004).



(Berka et al., 2005)

Figure 5: B-Alert System

Berka, the president of the B-Alert system, et al. (2005) conducted an experiment to determine if EEG could be used as an indicator of cognitive workload with the B-Alert System. The context of the experiment was a simulated Aegis command and control environment, a combat system with advanced, automatic detect-and-track, multi-function phased array radar. Five participants were trained as identification supervisors (IDSs). IDSs have one of the highest workloads of the 32 operators in a typical Aegis Combat Information Center (CIC). The IDSs were responsible for monitoring multiple data sources, detecting required actions, responding appropriately, and maintaining system status within predefined desirable parameters. Workload measures were calculated in real-time for each second of EEG by the B-Alert system. The B-Alert system used signal analysis techniques to decontaminate eye blinks, EMG data, amplifier saturation, and excursions related to movements. Post-hoc analysis identified the cognitive tasks from most difficult to least difficult were track selection-identification, alert-responses, hooking-tracks, and queries. High/extreme workload was detected approximately 100% of the time during high cognitive-load tasks such as selection-identification and alert-responses, 77% of the time for hooking-tracks, and 70% of the time for queries. The low false alarm rate of < 5% reflected that the workload gauge was not overly sensitive (Berka et al., 2005).

Although these results seem encouraging, currently no other known authors have reported such significant success with the B-Alert system and with the signal analysis techniques. A usability evaluation of a prototype of the Tactical Tomahawk Weapons Control System with the B-Alert system reflected that EEG correlated with expected workload when EEG measures were averaged across *all* nine participants. Individual participant EEG measures did not correlate strongly with cognitive workload, reflecting that the EEG was not a very sensitive indicator of cognitive workload. It was also found that EEG did not correlate with expert subjective ratings, but this may be due to insufficient data (Poythress et al., 2006).

The main benefits of EEG measurements are that they are continuous, cognitively unobtrusive, sensitive, moderately diagnostic, and relatively inexpensive. EEG provides a relatively continuous stream of data that can be identified and quantified on a second-by-second basis. Also, the decrease in size of electrodes and the development of wireless systems like B-Alert have led to minimal interference in primary task performance, unlike other functional neuroimaging techniques that require the subject to be completely immobile and are much more massive (Gevins & Smith, 2003). EEG is sensitive to changes in task complexity and task difficulty (Berka et al., 2004). The sensitivity of EEG is high and somewhat diagnostic as it can reflect subtle changes in attention, alertness, and cognitive workload (Berka et al., 2005; Gevins & Smith, 2003). Also, the technology required for EEG is fairly low in cost (Gevins & Smith, 2003).

However, the cost of the sensitivity is a poor signal-to-noise ratio. EEG can be easily contaminated by electrical signals of the eyes, muscles, heart, and external sources (De Waard, 1996). Also the considerable intra-subject and between subject variability makes it difficult to find consistent patterns of physiological change (De Waard, 1996). Signal analysis techniques prevent some of the signal-to-noise ratio problems; however, they must be rigorous and are time consuming. Required software and hardware is costly, and wired electrode systems may physically constrain participants. Advances have been made, but there is still a great deal of work to be conducted to accurately apply EEG measurements to cognitive workload assessment (Poythress et al., 2006).

Skin Measures

The most common skin measure used as an indicator of cognitive workload is based on the moment-to-moment sweat gland and related activities of the autonomic nervous system. When the body perspires, secreted positive and negative ions change the electrical properties of the skin. This phenomenon, originally called the psychogalvanic reflex (PGR), is currently known as the galvanic skin response (GSR), skin conductance response (SCR), or electrodermal response (EDR). In the remainder of this section, it will be referred to as EDR.

EDR can be performed passively or actively. In passive EDR, weak currents generated by the body itself are measured. Active EDR involves applying a small constant current through two electrodes on the skin such that a voltage develops across the electrodes. The skin acts as a variable resistor, and the effective resistance or conductance of the skin can be calculated by applying Ohm’s law (Allanson & Fairclough, 2004). The resulting EDR is expressed in terms of conduction or resistance, which are inversely related (De Waard, 1996).

EDR is frequently conducted actively when measuring cognitive workload. Electrodes are placed where sweat glands are most abundant such as the fingers, palm, forearm, and soles of the feet (Allanson & Fairclough, 2004; De Waard, 1996). In order to determine EDR, a baseline measure, which is typically an average of tonic EDR, is required. Tonic EDR is produced from everyday activities. It varies with psychological arousal and rises when the subject awakens and engages in cognitive effort, especially stress. In contrast, phasic EDR is the result of an external stimulus (De Waard, 1996). One-to-two seconds following an external stimulus, the skin conductance increases and peaks within five seconds (De Waard, 1996; McGraw-Hill Companies, 2007). The amplitude is dependent on the subjective impact of the stimulus, particularly with regard to its novelty and meaning. The wavelike increase in the curve is a fairly reliable indicator of the duration of information-processing and cognitive workload (Collet, Petit, Champely, & Dittmar, 2003).

The advantages of EDR are that it is low-cost and relatively simple to implement. The necessary hardware consists of a low-cost device called a galvanometer and inexpensive electrodes. An amplifier can also be added to increase the signal-to-noise ratio. Often only two electrodes are needed and can be affixed to the skin with adhesive tape.

The main disadvantages to EDR are the latency in response and the insensitivity to cognitive workload when compared to other physiological measures like heart rate variability and EEG. EDR is confounded by the many factors that impact the sympathetic nervous system and the sweat glands, which include temperature, respiration, humidity, age, gender, time of day, season, arousal, and emotions. Because EDR is related to activities of the sympathetic nervous system, all behavior (emotional and physical) can potentially change EDR (De Waard, 1996).

Situation Awareness Techniques

SA_T Calculation

$$SA_T = \frac{\text{Information Acquired}}{\text{Information Required}}$$

For Riese et al.’s work SA_T was based on three elements:

- Location (LOC) - Where is he? Knowing the position of entities or elements to a certain level of accuracy.
- Acquisition (ACQ) - What is he? Knowing the entity or element type to a particular level of acquisition (detect, classify, identify).
- State (STA) - How healthy is he? Knowing the mission-capable status of entities or elements.

The formula they used for the SA_T is below:

$$\text{Instantaneous } SA_T \text{ Score} = \frac{\sum_{i=1}^{n} c_i \left(W_L D_{Li} Loc_i + W_A D_{Ai} Acq_i + W_S Sta_i \right)}{\sum_{i=1}^{n} c_i \bullet \left(W_L + W_A + W_S \right)}$$

The SA_T result is between a 0 and 1, where 0 reflects a complete lack of useful SA information and 1 represents full knowledge of location, type, and state of enemy entities within the defined battlespace.

Team SA

Salmon et al (2006) evaluated seventeen SA measures against a set of human factors criteria for use in command, control, communication, computers and intelligence (C4i) environments. They concluded that current SA measurements are inadequate for use in assessing team SA, and recommended a multiple-measure approach. They identified three requirements for C4i SA measurement: ability to simultaneously measure SA at different locations, ability to measure SA in real-time, and ability to measure both individual and team/shared SA (Salmon, Stanton, N. Walker, G., & Green, 2006).

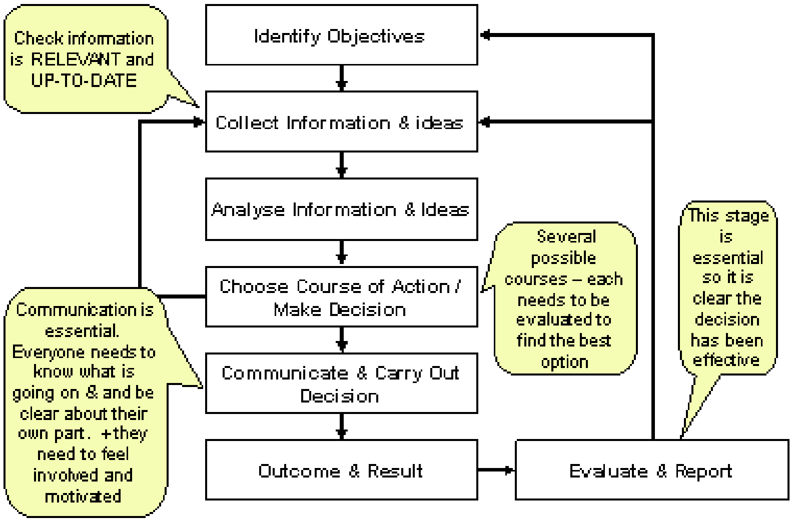
Measurement techniques for team SA have been researched and continue to be a subject of study. Some researchers have aggregated scores from validated SA measurement methods like SAGAT to obtain team SA scores; however, Gorman et al (2006) and others argue that combining individual SA results is not equivalent to team SA because it is missing interaction and critical activities such as coordination, collaboration, and information exchange (Entin & Entin, 2000; Gorman, Cooke, & Winner, 2006). Gorman et al. (2006) developed a measure of SA called the Coordinated Awareness of Situations by Teams (CAST) in which team SA is tested by the group’s reaction to unexpected events or “roadblocks.” They suggest that these “roadblock” situations are appropriate for determining shared SA because they require the team to coordinate and collaborate in order to circumvent the obstacle. The five steps to CAST involve identifying roadblocks, documenting primary perceptions, documenting secondary perceptions, documenting coordinated perceptions, and documenting coordinated actions (Gorman et al., 2006).

Another team SA technique, Situational Awareness Linked Instances Adapted to Novel Tasks (SALIANT), is similar to CAST. It involves how the team responds to problems, but requires an a priori script and structure. SALIANT includes five phases: 1) identify team SA behaviors (e.g., demonstrated awareness of surrounding environment, recognized problems, anticipated a need for action, demonstrated knowledge of tasks, demonstrated awareness of information) 2) develop scenarios, 3) define acceptable responses, 4) write a script, and 5) create a structured form with columns for scenarios and responses (Gawron, 2000).

Related to team SA is the concept of distributed situation awareness (DSA) (Stanton et al., 2006). Unlike definitions of team SA that examine shared SA, DSA focuses on the system as a whole and consists of the complimentary SA of agents (humans and technology) within the system. Although DSA captures the content of system SA it does not have a technique for assessing the quality of that SA, other than task performance and SME judgment.

Decision Making

In general, the components of decision making, for both system and humans, include identifying objectives, collecting information and ideas, analyzing the information and ideas, choosing a course of action/making a decision, and communicating and carrying out the decision. An outcome results from the action taken and, optimally, there is an evaluation or report reflecting the effectiveness of the decision (Figure 6).

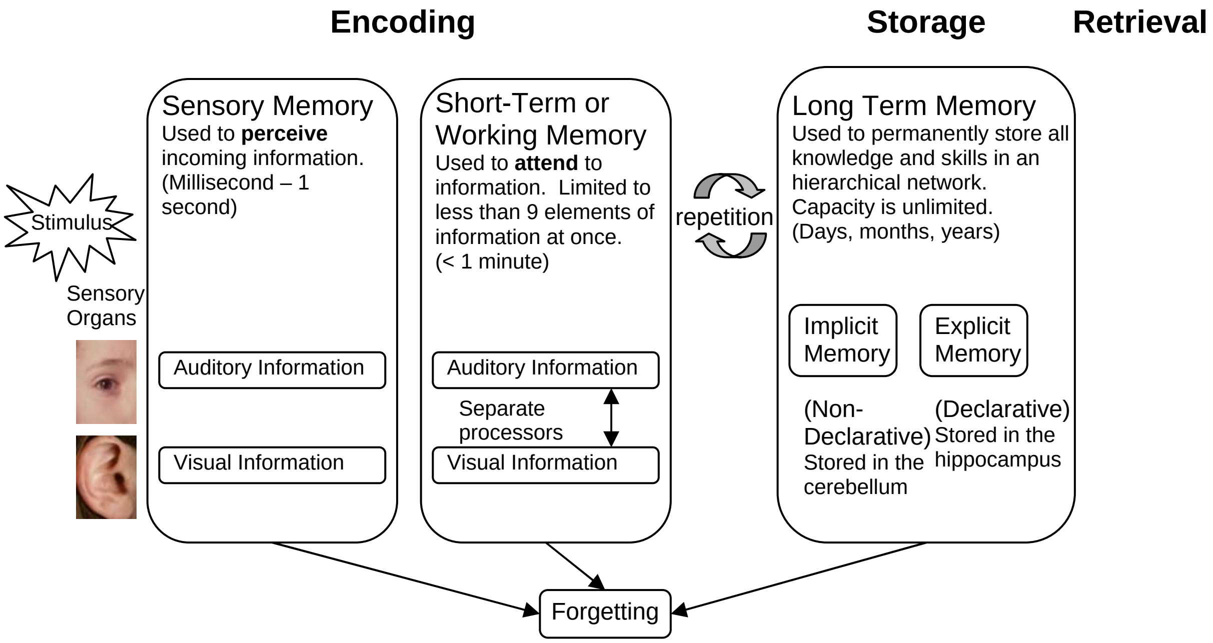

Check information is RELEVANT and UP-TO-DATE
Identify Objectives
Collect Information & ideas
Analyse Information & Ideas
Choose Course of Action / Make Decision
Communicate & Carry Out Decision
Outcome & Result
Evaluate & Report
Communication is essential. Everyone needs to know what is going on & and be clear about their own part. +they need to feel involved and motivated
Several possible courses – each needs to be evaluated to find the best option
This stage is essential so it is clear the decision has been effective

(Finntrack,)

Figure 6: General components of decision making

Decision making depends in part on the system providing the “right” information at the “right” time to the “right” person. In order to achieve this goal, an Internet-like system called the Global Information Grid (GIG) is in the conceptual design phase for C2. The Department of Defense defines the GIG as a “globally interconnected, end-to-end set of information capabilities, associated processing, storing, disseminating, and managing information.” The GIG will be a repository for military information with the goal of providing information superiority over adversaries. A significant risk of the GIG is inundating warfighters with information and providing noisy data that distracts from mission critical data (Bass & Baldwin, 2007). Since net-centric, GIG-type environments are slated for the future, many papers have been written about the GIG and its potential implications. Bass and Baldwin (2007) proposed some rules to direct GIG information flow because information presented at the wrong time, level of detail, and/or lacking proper analysis and interpretation could have devastating effects. Their basic solution is to limit data access to those with authorization and limit data automatically sent to all users (Bass & Baldwin, 2007). The goal is to present a manageable amount of information to warfighters that pertains to them, but also provides context.

Cognitive aspects of decision making begin when the decision maker attends to information in his/her environment. The brain selects data for cognitive processing and encodes data for storage and later retrieval to support decision making (Figure 7). Encoding, translating stimuli to internal (mental) representations, is the longest component of reasoning and decision making taking approximately 45% of the overall time.


Encoding
Storage
Retrieval
Stimulus
Sensory Organs
Sensory Memory
Used to **perceive** incoming information. (Millisecond – 1 second)
Auditory Information
Visual Information
Short-Term or Working Memory
Used to **attend** to information. Limited to less than 9 elements of information at once. (< 1 minute)
Auditory Information
Separate processors
Visual Information
repetition
Long Term Memory
Used to permanently store all knowledge and skills in an hierarchical network. Capacity is unlimited. (Days, months, years)
Implicit Memory
Explicit Memory
(Non-Declarative) Stored in the cerebellum
(Declarative) Stored in the hippocampus
Forgetting

Composed from http://www.scientificjournals.org/journals2007/articles/graphics/1038.gif, http://thebrain.mcgill.ca/flash/i/i_07/i_07_p/i_07_p_tra/i_07_p_tra_2a%20copy.jpg, and http://education.arts.unsw.edu.au/staff/sweller/clt/images/CLT_NET_Aug_97_HTML6.gif

Figure 7: Cognitive Aspects of Decision Making

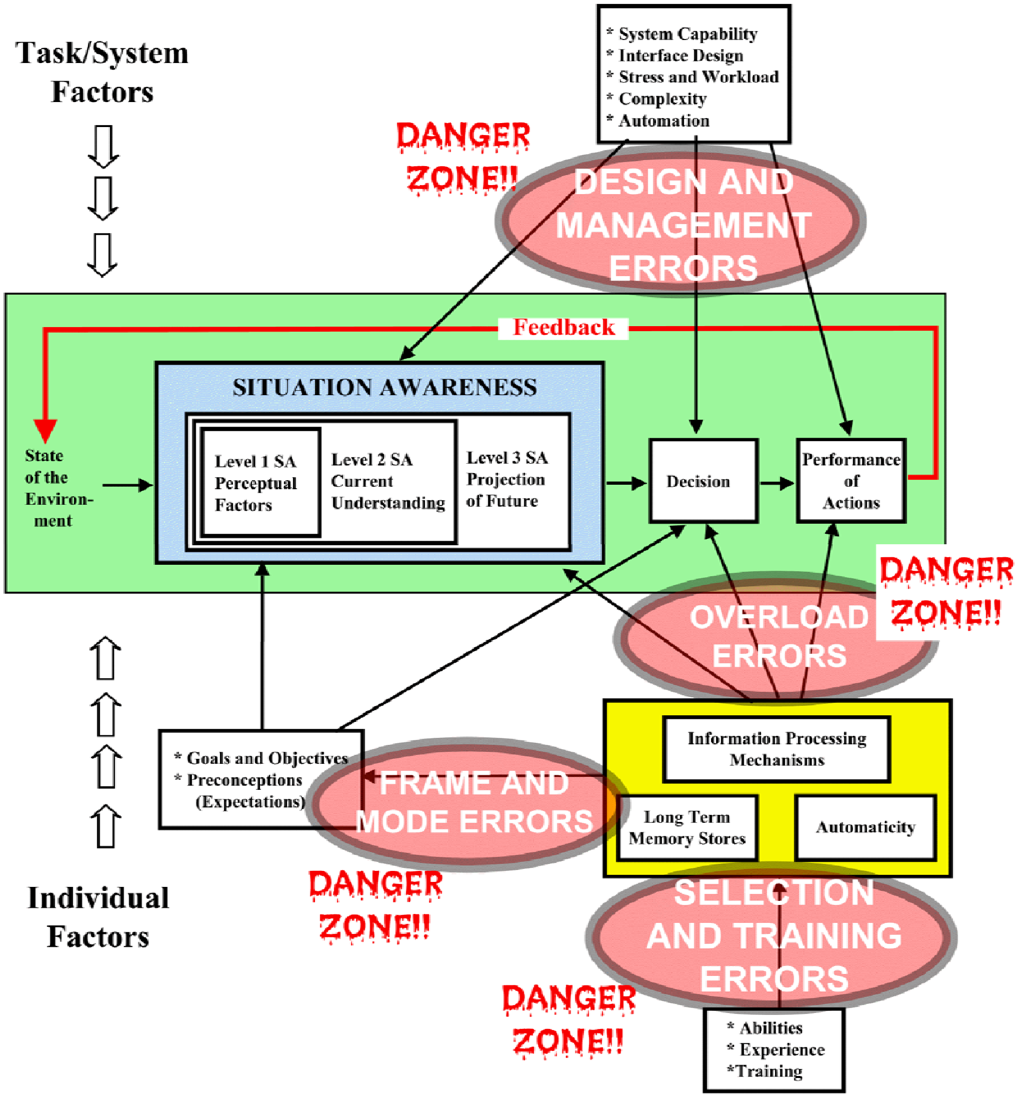
Encoding words takes longer and requires more workload than encoding schematic pictures, so reducing the amount of text will lead to faster decision making. Time-critical decision making displays should aim to decrease encoding time, for example by increasing the intuitiveness of symbology and tasking (Azuma, Daily, & Furmanski, 2006).

The encoding, storage, and retrieval process reflects the close connections between cognitive workload, situation awareness, and decision making. Part of encoding involves perceiving and comprehending, which are critical components of workload and situation awareness. Cognitive workload is related to the storage process because if encoding takes longer or if the information is encoded incorrectly, it may be because the information is more difficult and requires more cognitive resources to process. SA is also an integral part of the storage process because it directly involves perception and comprehension. Typically, cognitive workload and SA are inversely proportionate: a decrease in workload often results in an increase in SA and vice versa. Hence, if the storage process is improved, there is a higher likelihood that cognitive workload will decrease, situation awareness will increase, and decision making will improve as long as the information is relevant to the decision makers' decision.

The projection aspect of SA provides the foundation for decision making. Humans are adept at detecting and remembering patterns. When they notice trends, they begin to extrapolate from the trends to anticipate what will happen next. Series of trends over time result in the development of mental models. This process is called sensemaking. The Command and Control Research Program's Sensemaking Symposium Final Report defines sensemaking as, "the process of creating situation awareness in situations of uncertainty" (Leedom, 2001). Using a data/frame theory as an analogy, data are associated to form a hypothesis called a frame. Preserving and elaborating the frame are like Piaget's assimiliation and reframing is like Piaget's accommodation. Once anchors are established to generate a useful frame, the frame can be evaluated, reframed, elaborated, or compared with alternative frames (G. Klein, Moon, & Hoffman, 2006b). Sensemaking is similar to Endsley's model of situation awareness except instead of a knowledge state that's achieved, sensemaking is the process of getting to the outcome, the strategies, and the obstacles. "Sensemaking is a motivated, continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories, and act effectively" (G. Klein, Moon, & Hoffman, 2006a). Decision makers apply their mental models to the tasks they conduct. Figure 8 provides a model of the relationships between SA and decision making (DM) aspects. In addition, it shows the influence of both the technical/system environment as well as the individual. Finally, it highlights the many areas that can produce errors.

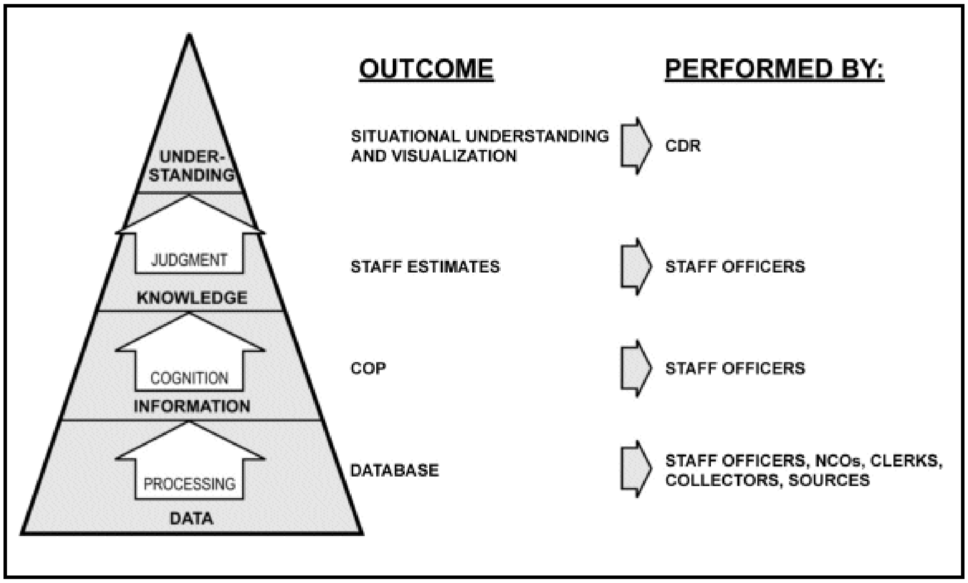
There are several decision making modeling techniques to emulate human decision making. For instance, speeding up decision time provides a tactical advantage in line with John Boyd's high-level sequential model of decision making (consisting of Observation, Orientation, Decision, and Action) called the OODA loop, in which the basic C2 strategy is to execute your OODA loop faster than your opponent (Azuma et al., 2006). Several cognitive models in addition to John Boyd's OODA loop have been developed, for example, ACT-R, EPIC, GOMSL, Soar, IMPRINT, and Bayesian networks (Endsley, Hoffman, Kaber, & Roth, 2007).

Cognitive models tend to generalize decision making; and in uncertain and less familiar situations like complex C2 environments, individual differences, particularly experience, become more apparent (Lamar, 2006). Experience level and prior exposure to similar tasks are discriminating factors between individuals. With novel incoming stimuli, individuals are biased by their past experience, values, and repercussions of the decision (i.e., the cost-benefit analysis) when choosing a course of action (Lamar, 2006). Rasmussen and Reason described experience using a skill-, rule-, and knowledge-based (SRK) classification. Individuals at a skill level have a great deal of experience with the task and the environment and behave virtually automatically. Rule based behavior involves less experience than the skill level and follows an "if, then" process. Knowledge based behavior is unskilled and requires more effort because less memory can be leveraged. Figure 9 below shows an application of SRK based classification to C2.



(Smith, 2007)

Figure 8: Relationship of SA and DM aspects



(Department of the Army, Apr 2003)

Figure 9: SRK Classification of C2

Another individual difference that affects decision making is physiology. Some individuals are innately better and quicker at decision making than others, but aspects of decision making can be learned. For instance, taxi drivers who memorized a map had a different degree of hippocampal volume, a space involved with learning and memory, when compared to those who had not (Lamar, 2006). Expertise leads to differences in function and structure of brain regions required for decision making. Thus, decision making is influenced by both nature and nurture. Another individual difference is personality, but research is ambiguous as to the effect of personality on decision making (Koterba, 2004). Another significant individual difference in decision making is confidence (Sniezek, 1992). Confidence can have critical consequences as disasters frequently result when decision makers are confidently wrong. Conversely, if decision makers have reached an accurate decision and don't have the confidence in their decision to follow-through, the potential benefits would be lost. (Sniezek, 1992)

Most individual difference effects are superseded by trained decision making strategies and short-cuts in time-critical environments like C2 (Koterba, 2004; Lehner, Seyed-Solorforough, O'Connor, Sak, & Mullin, 1997). C2 decision makers often have to follow protocol and procedures they have been trained to execute (Lehner et al., 1997). When C2 decision makers do not have protocol or procedures to follow, they apply rules of thumb or "error-prone heuristics" that result in "good enough" versus exact results obtained from rational decision making outputs (Albers, 1996). Decision makers are especially prone to applying heuristic decision processing under stressful conditions (Lehner et al., 1997). This is a natural tendency because humans have limitations on their capacity to process information, and cope by grouping information and applying mental shortcuts. Kahneman and Tversky pioneered research in this area and began compiling individual cognitive and personal biases in decision making that currently are part of a long list. Some examples relevant to C2 are below:

- Availability Bias: Tendency to overestimate usual or easy to remember events (Lehner et al., 1997)

- Recency Bias: Emphasize recent information; tend to forget or ignore older data (??, 2007)
- Short evidence search: Accept the first alternative that seems credible (??, 2007)
- Source credibility bias: Accept or reject information depending on personal relations with source (??, 2007)
- Ascription of causality: Over generalize correlation as causation (??, 2007)
- Fundamental Attribution Error: View successes as results of talent and failures due to external factors or bad luck, attribute the success of others to good luck and failures to their mistakes (??, 2007; Lehner et al., 1997)
- Hindsight Bias: After an answer is known, people suggest that they knew all along, when they were unclear at the onset (Lehner et al., 1997)
- Choice-supportive bias: Disregard negative aspects of chosen and highlight negative aspects of rejected options to justify choices (??, 2007)
- Inertia: Unwilling to change past thought patterns for new situations (??, 2007)
- Role Fulfillment (Self Fulfilling Prophecy): Causing what is believed will happen to happen (??, 2007)
- Group Think: Willingness to conform to popular group opinion (??, 2007)
- Underestimating uncertainty and the illusion of control: Employ an overly optimistic perception of personal control and not place enough emphasis on the degree of uncertainty (??, 2007)
- Automation Bias: Over reliance on automation (Cummings, Bruni, Mercier, & Mitchell, 2007)
- Anchoring and adjustment: Initial information influences choice (Tatarka, 2002)
- Confirmation bias: find evidence that supports preconceived conclusion and disregarding evidence that contradicts conclusion (Lehner et al., 1997; Tatarka, 2002)

An article from the Military Intelligence Professional Bulletin reports that a large amount of anecdotal evidence suggests the two most dangerous and common biases in military C2-like operations are anchoring and adjustment and confirmation biases (Tatarka, 2002). Once an intelligence analyst has anchored (anchoring and adjustment) on an enemy course of action, they seek evidence that confirms their decision and disregard conflicting information (confirmation bias). The authors suggest that military doctrine promotes this bias because of short time constraints. The risk is “cognitive tunnel vision”, which is emphasized in high stress situations like C2, and could lead to devastating effects (Tatarka, 2002). Another potentially dangerous and common heuristic in C2 is automation bias. As technology becomes more prevalent and can provide automated solutions, operators may over rely on the automation, become complacent, and may experience a loss of situation awareness (Cummings et al., 2007). These cognitive biases are highly relevant to consider in decision making tasks because they become increasingly resistant to preventative techniques such as, training, decision support tools, and devil’s advocate approaches, in unfamiliar and stressful environments like complex C2 (Lehner et al., 1997; Tatarka, 2002). The heuristics and biases can be better understood by applying techniques like those in understanding information requirements; for example, cognitive task analysis, cognitive work analysis, etc.

CTA Methodologies

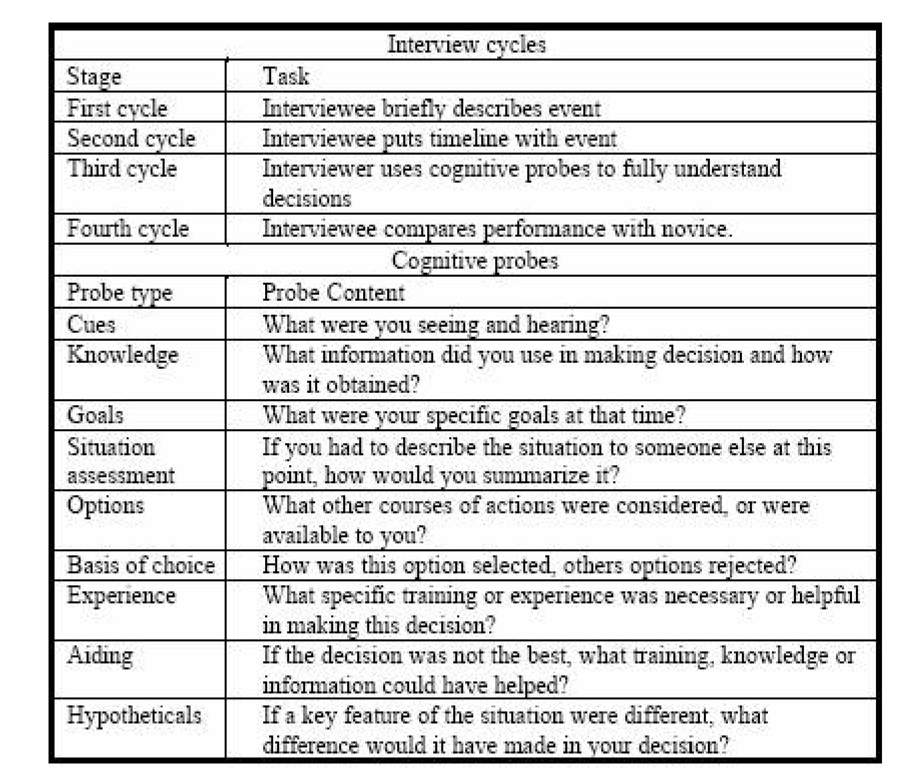
A number of methods exist to obtain information requirements and decision rationale from the decision maker, the most common being cognitive task analysis (CTA). A variety of CTA methodologies have been developed that differ in approach, structure, emphasis, and resource requirements, but all include some sort of knowledge elicitation, analysis, and knowledge representation (Federal Aviation Administration Human Factors Division, 1999; Militello & Hutton, 1998). Often the knowledge elicited is not measured quantitatively; however, aspects that can be measured quantitatively are number or type of ideas considered and the number or type of consequences considered. The analysis and knowledge representation are the significant parts of the CTA as they can be used to improve processes or systems. If the system and process are in line with the decision makers' mental model derived from conducting a CTA, there is a higher likelihood that the decision maker will be able to use the system more effectively and efficiently to make a decision.

Three examples of CTA are: 1) Precursor, Action, Results, and Interpretation (PARI) method 2) Critical Decision Method (CDM), and 3) Conceptual Graph Analysis (CGA). In the PARI method, subject matter experts identify troubleshooting use cases to elicit system knowledge (how the system works), procedural knowledge (how to perform problem solving procedures), and strategic knowledge (knowing what to do and when to do it) from other subject matter experts (Federal Aviation Administration Human Factors Division, 1999; Jonassen, Tessmer, & Hannum, 1998). "PARI attempts to identify each *A*ction (or decision) that the problem solver performs, the *P*recursor (or Prerequisite) to that action, the *R*esult of that action, and an expert's *I*nterpretation of the Results of that Action" (Jonassen et al., 1998). The PARI technique, developed by Hall, Gott, and Pokorny in 1995, was originally designed for the Air Force to design intelligent tutoring systems. It involves a structured interview during which pairs of subject matter experts probe each other under realistic conditions. The interviews are held during and after troubleshooting has occurred with an emphasis on reasoning used in making decisions. PARI products include flowcharts, annotated equipment schematics, and tree structures (Federal Aviation Administration Human Factors Division, 1999).

An advantage of PARI is that it thoroughly exposes how subject matter experts deal with systems by identifying the technicalities of how the system works (technology focused), how to perform problem solving procedures (human-system interface), and knowing what to do about the problem (cognitive or decision making rationale). PARI is especially strong in revealing troubleshooting and analyzing problem solving techniques that can be beneficial for training. The disadvantage of PARI is that it may focus too much on specific troubleshooting and it relies heavily on subject matter expertise. Since PARI consists of subject matter experts interviewing each other, there is a risk that some details may be left out because they may be assumed to be included or considered trivial.

CDM, based on Flanagan's critical incident technique developed in 1954, and formalized by Klein in 1993, is a series of semi-structured interviews of subject matter experts that focuses on critical, non-routine incidents requiring skilled judgment (G. A. Klein, Calderwood, & Macgregor, 1989). The interview is considered semi-structured because it is in between an ongoing verbal protocol where the decision maker "thinks aloud" and a completely structured interview (G. A. Klein et al., 1989). The theory behind CDM is that probing subject matter experts about difficult activities results in the "richest source of data" to understand decision making of highly skilled personnel as the information gleaned is expertise, not formalized protocol (G. A. Klein et al., 1989). When CDM is conducted, subject matter experts recount a difficult incident and the interviewer probes to distinguish decision points, critical cues, cognitive strategies, etc. (Table 1 provides information on interview cycles and example cognitive probes.) (Federal Aviation Administration Human Factors Division, 1999) p.126.

Table 1: CDM Interview Cycles and Cognitive Probes



Interview cycles	
Stage	Task
First cycle	Interviewee briefly describes event
Second cycle	Interviewee puts timeline with event
Third cycle	Interviewer uses cognitive probes to fully understand decisions
Fourth cycle	Interviewee compares performance with novice.
Cognitive probes	
Probe type	Probe Content
Cues	What were you seeing and hearing?
Knowledge	What information did you use in making decision and how was it obtained?
Goals	What were your specific goals at that time?
Situation assessment	If you had to describe the situation to someone else at this point, how would you summarize it?
Options	What other courses of actions were considered, or were available to you?
Basis of choice	How was this option selected, others options rejected?
Experience	What specific training or experience was necessary or helpful in making this decision?
Aiding	If the decision was not the best, what training, knowledge or information could have helped?
Hypotheticals	If a key feature of the situation were different, what difference would it have made in your decision?

A variety of CDM products can be produced; one of the most common is a narrative account (Federal Aviation Administration Human Factors Division, 1999). Another product is a cognitive requirements table that includes cognitive demands of the task and pertinent contextual information. CDM results are usually used to develop system design recommendations or training (Federal Aviation Administration Human Factors Division, 1999).

CDM was implemented in a C2 decision making study of anti-air warfare operators on a U.S. Navy AEGIS cruiser to investigate decision maker strategies. Results reflected that the feature matching strategy, involving recognition of a typical class of situation, was the most used strategy (87% of diagnostic strategies). Story building was also used (12% of diagnostic strategies) where the situation was novel or where the decision maker builds a story from seemingly disparate pieces of information to develop a coherent explanation of the situation. Decision makers did not evaluate 75% of the decisions that they implemented, and considered and compared multiple options in only 4% of the cases. In the 4% of cases where multiple options were considered, they were not the most critical decision points. When decision makers did not understand a situation, they prepared for the worst case scenario probably to avoid risk (Kaempf, Klein, Thordsen, & Wolf, 1996).

The advantage of CDM is that it reveals expertise and understanding of objectives that would not otherwise be illuminated. The semi-structured organization provides flexibility to the decision maker to discuss aspects that might not have been specified a priori. It has also been used to in complex C2 environments to determine decision making strategies. The interview cycle approach expands the attributes of information collected, and also increases the time and resources required to conduct CDM. A disadvantage of CDM is that it is subjective and reflective on the decision maker's own strategies and basis for decisions (G. A. Klein et al., 1989). Another disadvantage is that the critical event chosen may be very atypical or rare. Finally, since CDM is less structured, it is more difficult to interpret and analyze the results.

CGA was developed in 1992 by Gordon and Gill and involves generating a visualization of conceptual knowledge structures to conduct CTA (Federal Aviation Administration Human Factors Division, 1999). A CGA consists of a formal and detailed collection of nodes (which can be goals, actions or events), relations, and questions (Federal Aviation Administration Human Factors Division, 1999; Jonassen et al., 1998). Nodes are connected via arcs, which portray the relationship between nodes. The CGA process begins by exploring any pre-existing documentation related to the task to be analyzed. Then, a process called free generation is implemented in which an SME leverages the existing documentation and adds task information requirements. The information is then compiled and visually presented as a draft conceptual graph. Any gaps in the representation are constructed into detailed questions. If there are still gaps after questions are asked, information is filled in from observations (Federal Aviation Administration Human Factors Division, 1999). The last step is validating the conceptual graph by having an expert perform the task and check for incorrect or missing information (Jonassen et al., 1998).

An advantage to CGA is that it provides a visual depiction of internal knowledge like a concept map. Clarifying the linkages between concepts causes the interviewer to closely investigate the conceptual relationships that might not be examined through other CTA techniques. Another advantage is the detailed approach affords a systematic process with more structure than other CTA methods. The structure also yields "specific yet comprehensive" questions. In addition, a variety of automated software tools exist to assist in developing conceptual graphs like COG-C. A disadvantage is the CGA nodes and arcs take time to learn, and a CGA is difficult to develop while an unstructured interview is taking place. Also, while CGA describes concepts well, it is weak at capturing procedural knowledge (Jonassen et al., 1998).

Despite the contrast between CTA techniques, they are useful in revealing C2 decision maker rationale. The PARI technique is considered a traditional cognitive task analysis technique, CDM is considered activity-based analysis, and CGA is considered subject matter/content analysis; however, all reveal information that could improve C2 processes, or be used to evaluate C2 decision making. Most of these techniques focus on deviations from standard operating procedures and preplanned responses where C2 decision making and expertise can be exposed (Jonassen et al., 1998). In addition, the various levels of structure in CTA methodologies parallels the levels of structure in various aspects of C2 (G. A. Klein et al., 1989). Another positive aspect is that many of the CTA techniques are conducted retrospectively, which is important in C2 because they are less intrusive (G. A. Klein et al., 1989). The caveat to CTAs are that "no well-established metrics exist" for evaluating CTAs, and it is difficult to evaluate differences between CTA methods (Militello & Hutton, 1998). This is partially because it is unknown what information is lost versus gained in comparison to other techniques and also because interviewees and individuals provide different information each. Also, CTAs can be very resource intensive. Because individual differences impact how much information individuals are willing to provide and respond, it is difficult to assess the reliability and validity of CTA methods. Also, another caveat is that no advanced techniques for team CTA have been developed (Militello & Hutton, 1998).

References

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics, 36,* 623-636.

Albers, M. J. (1996). Decision making: A missing facet of effectivfe documentation. *ACM Special Interest Group for Design of Communication: Proceedings of the 14th Annual International Conference on Systems Documentation: Marshaling New Technological Forces: Building a Corporate, Academic, and User-Oriented Triangle,* , 57-65.

Allanson, J., & Fairclough, S. H. (2004). A research agenda for physiological computing. *Interacting with Computers, 16*(5), 857-878.

Azuma, R., Daily, M., & Furmanski, C. (2006). A review of time critical decision making models and human cognitive processes. *Aerospace Conference, 2006 Institute of Electrical and Electronics Engineers, Inc.,*

Basar-Eroglu, C., & Demiralp, T. (Jan 2001). Event-related theta oscillations: An integrative and comparative approach in the human and animal brain. *International Journal of Psychophysiology, 39*(2-3), 167-195.

Bass, S. D., & Baldwin, R. O. (2007). A model for managing decision-making information in the GIG-enabled battlespace. *Air and Space Power Journal,* , 100-108.

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovis, M. M., Davis, G., Lumicao, M. N., et al. (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.

Berka, C., Levendowski, D. J., Ramsey, C. K., Davis, G., Lumicao, M. N., Stanney, K., et al. (2005). Evaluation of an EEg-workload model in the aegis simulation environment. Paper presented at the *, 5797* 90-99.

Boehm-Davis, D. A., Gray, W. D., & Schoelles, M. J. (2000). The eye blink as a physiological indicator of cognitive workload. *Proceedings of the IEA 2000/HFES 2000 Conference,*

Collet, C., Petit, C., Champely, S., & Dittmar, A. (2003). Assessing workload through physiological measurements in bus drivers using and automated system during docking. *Human Factors, 45*(4), 539-548.

Colman, A. M. (2001). A dictionary of psychology: P300. ()Oxford University Press.

Cummings, M. L., Bruni, S., Mercier, S., & Mitchell, P. J. (2007). Automation architecture for single operator-multiple UAV command and control. *The International C2 Journal: Special Issue- Decision Support for Network-Centric Command and Control, 1*(2), 1-24.

Dantest Medical System. *What is heart rate variability (HRV) analysis?* Retrieved 12/4, 2008, from http://www.dantest.com/introduction_what_is_hrv.htm

De Waard, D. (1996). The measurement of drivers' mental workload. (PhD thesis, University of Groningen).

Department of Kinesiology University of Waterloo. *Parietal lobe.* Retrieved 1/13, 2008, from http://ahsmail.uwaterloo.ca/kin356/dorsal/parietal.jpg

Department of the Army. (Apr 2003). *Field manual 3-21.21: The stryker brigade combat team infantry battalion*

Endsley, M. R., Hoffman, R., Kaber, D., & Roth, E. (2007). Cognitive engineering and decision making: An overview and future course. *Journal of Cognitive Engineering and Decision Making, 1*(1), 1-21.

Entin, E. B., & Entin, E. E. (2000). Assessing team situation awareness in simulated military missions. (1) 73-76.

Federal Aviation Administration Human Factors Division. (1999). *Department of defense handbook: Human engineering program process and procedures* No. MIL-HDBK-46855A)

Finntrack. *Decision making process.* Retrieved 2/12, 2008, from http://www.finntrack.com/hnc_hnd/bus-decision.htm

Gawron, V. J. (2000). *Human performance measures handbook*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science, 4*(1-2), 113-131.

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics, 49*(12-13), 1312-1325.

Hannula, M., Koskelo, J., Huttenen, K., Sorri, M., & Leino, T. (2007). Artificial neural network analysis of heart rate under cognitive load in a flight simulator. Paper presented at the

Harris, B. *A revolution in neuroscience: Tuning the brain.* Retrieved 1/13, 2008, from http://www.centerpointe.com/about/articles_research.php

Jonassen, D. H., Tessmer, M., & Hannum, W. H. (1998). *Task analysis methods for instructional design*

Kaempf, G. L., Klein, G., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors, 38*(2), 220-231.

Klein, G., Moon, B., & Hoffman, R. R. (2006a). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems, 21*(4), 70-73.

Klein, G., Moon, B., & Hoffman, R. R. (2006b). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems, 21*(5), 88-92.

Klein, G. A., Calderwood, R., & Macgregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics, 19*(3), 462-472.

Koterba, N. T. (2004). *APL internal report: The effect of personality on decision making*

Lamar, M. (2006). Neuroscience and decision making.

Leedom, D. K. (2001). *Sensemaking symposium final report*Command and Control Research Program.

Lehner, P., Seyed-Solorforough, M., O'Connor, M. F., Sak, S., & Mullin, T. (1997). Cognitive biases and time stress in decision making. *IEEE Transactions on Systems, Man, and Cybernetics- Part A: Systems and Humans, 27*(5), 698-703.

McGraw-Hill Companies, I. (2007). *Science and technology encyclopedia, 5th edition: Electrodermal response.* Retrieved 5/18, 2007, from http://www.answers.com/topic/galvanic-skin-response

Militello, L. G., & Hutton, R. J. B. (1998). Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 41*(11), 1618-1641.

Nickel, P., & Nachreiner, F. (2003). Sensitivity and diagnosticity of the 0.1-hz component of heart rate variability as an indicator of mental workload. *Human Factors, 45*(4), 575-590.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. Boff, L. Kaufman & J. Thomas (Eds.), *Handbook of perception and performance* (vol. 2 ed.,). New York: Wiley.

Poole, A. (2004). Eye movements.

Poythress, M., Russell, C., Siegel, S., Tremoulet, P. D., Craven, P., Berka, C., et al. (2006). Correlation between expected workload and EEG indices of cognitive workload and task engagement. *2nd Annual AugCog International Conference*, San Fransisco, CA. 32-44.

Rowe, D. W., Silbert, J., & Irwin, D. (1998). Heart rate variability: Indicator of user state as an aid to human-computer interaction. Paper presented at the 480-487.

Salmon, P., Stanton, N. Walker, G., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics, 37*, 225-238.

Smith, D. J. (2007). *Situation(al) awareness in effective command and control.* Retrieved 1/3, 2008, from http://www.smithsrisca.demon.co.uk/situational-awareness.html

Sniezek, J. (1992). Groups under uncertainty: An examination of confidence in group decision making. *Organizational Behavior and Human Decision Making Processes*, (52), 124-155.

Tatarka, C. J. (2002, Overcoming biases in military problem analysis and decision-making. [Electronic version]. *Military Intelligence Professional Bulletin*, (Jan-March, 2002)

Tattersall, A. J., & Hocky, G. R. J. (1995). Level of operator control and changes in heart rate variability during simulated flight maintenance. *Human Factors, 37*(4), 682-698.

Tsai, Y., Viirre, E., Strychacz, C., Chase, B., & Jung, T. P. (2007). Task performance and eye activity: Predicting behavior relating to cognitive workload. *Aviation, Space, and Environmental Medicine: Operational Applications of Cognitive Performance Enhancement Technologies,*

Van Orden, K. F. (2000). *Real-time workload assessment and management strategies for command and control watchstations: Preliminary findings.* Unpublished manuscript. Retrieved March 16, 2006, from http://www.dtic.mil/matris/sbir/sbir011/Navy89b.doc

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors, 1*, 111-121.

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics, 41*(5), 656-669.

Wickens, C. D., Mavor, A. S., & McGee, J. P. (Eds.). (1997). *Flight to the future: Human factors in air traffic control*National Academy Press.