2011

Cognitively Sensitive User Interface for Command and Control Applications

Michael James Findler
Wright State University

Cognitively Sensitive User Interface for Command and Control Applications

A dissertation in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

By

MICHAEL JAMES FINDLER
B.S., Loyola Marymount University, 1981
M.S., University of Dayton, 1987
M.S.E., Wright State University, 2002

2011
Wright State University

WRIGHT STATE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

March 16, 2011

     I HEARBY RECOMMEND THAT THE DISSERTATION PROPOSAL PREPARED UNDER MY SUPERVISION BY <u>Michael J. Findler</u> ENTITLED <u>Cognitively Sensitive User Interface for Command and Control Applications</u> BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF <u>Doctor of Philosophy</u>.

_____
Dr. S. Narayanan, Ph.D.
Dissertation Director
Ph.D. Program

_____
Ramana Grandhi, Ph.D.
Director, Ph.D. in Engineering Program

_____
Andrew Hsu, Ph.D.
Dean, School of Graduate Studies

Committee on Final Examination

_____
Dr. S. Narayanan, Ph.D.

_____
Mateen Rizki, Ph.D.

_____
Yan Liu, Ph.D.

_____
Joseph Litko, Ph.D.

_____
Misty Blue-Terry, Ph.D.

ABSTRACT

Findler, Michael James. Ph.D. Department of Biomedical, Industrial, and Human Factors Engineering, Wright State University, 2011. Cognitively Sensitive User Interface for Command and Control Applications.

In many complex systems, such as nuclear power plants, etc., human decision makers are required to make critical decisions in a time-pressured environment. Typically, most of these applications are dynamic and uncertain and require humans making supervisory control decisions through monitoring, re-planning, troubleshooting, and control. Due to the critical nature of decision making, human operators are responsible for the safe and efficient operation of these applications. Delays and failures in making decisions, in these applications, are often expensive in terms of money, system performance, and may even cost human lives. In a system where human supervisors control computerized processes, the human must work seamlessly with the computerized system in achieving overall system objectives. Research on human-centered automation in aviation, satellite ground control, and nuclear power plant control has resulted in broad guidelines on system design involving human and computerized processes in supervisory control. However, problems remain, such as increased human error, lack of situational awareness, and opacity from poorly automated systems, particularly in scenarios where human operators must make decisions in time-pressured planning.

A key aspect of overcoming these problems is to effectively couple human decision makers with the computerized systems through user interface design. Context-

free research, i.e., research on 'best' menu structure or set of colors for display screens, is not typically useful in enhancing operator or system performance in which operators are well motivated. Research on interface design for complex systems for example, demonstrates that semantic issues such as display content, level of abstraction, visual momentum across windows, etc., quickly dominate the effects of the primarily syntactic aspects of the human-computer interface, e.g., color, selection style, etc.

While there are broad guidelines for display or user interface design, creating effective human-computer interfaces for complex, dynamic systems control is challenging. *Ad hoc* approaches which consider the human as an afterthought are limiting. This research proposes a systematic approach to human / computer interface design that focuses on both the semantic and syntactic aspects of display design in the context of human-in-the-loop supervisory control of intelligent, autonomous multi-agent simulated unmanned aerial vehicles (UAVs). A systematic way to understand what needs to be displayed, how it should be displayed, and how the integrated system needs to be assessed is outlined through a combination of concepts from naturalistic decision making, semiotic analysis, and situational awareness literature. A new sprocket-based design was designed and evaluated in this research.

For the practical designer, this research developed a systematic, iterative design process: design using cognitive sensitive principles, test the new interface in a laboratory situation; bring in subject matter experts to examine the interface in isolation; and finally, incorporate the resulting feedback into a full-size simulation. At each one of these steps, the operator, the engineer and the designer reexamined the results. The goal is to present

to the operator a more complete feel for the complex system, and one that can evolve with the operator's experience. Individual user interface components were empirically examined before acceptance into the integrated global design.

Two laboratory-based controlled experiments were designed and executed: a static decision-aiding display showing relative merits for alternate mission routes and a dynamic system monitoring display showing each UAVs health status. Following the success of the first two experiments, the results were presented to domain subject matter experts, at which time further refinements were suggested. Finally, this was followed up by a hands-on demonstration of the displays embedded within a simulator testbed.

This research provided useful insights into human decision making in complex systems. It examined multiple user interfaces: tabular text-based, graphical bar charts, analog gauges and Visual Thinking widgets. Results are promising. The Visual Thinking Sprocket display was significantly better as a decision aid and as a system monitoring display. Responses were faster and more accurate using the Visual Thinking Sprocket. This research presents a practical framework that can be systematically applied to the designing and producing of cognitively sensitive displays in complex, dynamic systems control.

Acknowledgements

First, I would like to thank my lovely wife, Teri W. Miller, for her patience and editing "prowess". My daughters, Teagan and Moreta, they know no other life than daddy being in school.

My parents believe in me, even after so many years. My brothers and sisters, Marie, Daniel, Mark, Steven, Christopher, Paul and Angelica for their believe that I was smart enough – even when I didn't think so.

My Committee: Drs. S. Narayanan, Mateen Rizki, Joseph Litko, Yan Liu and Misty Blue-Terry. I would like to especially thank Dr. S. Narayanan for taking a chance on me. Thank you for pointing me in the right direction – over and over again. I will pay it forward to those I am to help.

Sytronics, Inc. for their support, especially the late Dr. William Marshak. Bill, you still have a perfect record in graduating those you help.

My labmates for their endless support and camaraderie.

Table of Contents

Table of Figures

Table of Tables

Acronyms

AAA	Anti-Aircraft Artillery
AD	Air Defense
DHS	Department of Homeland Security
DSI	Decision Support Interface
DSS	Decision Support System
EW	Electronic Warfare
FoM	Figure(s) of Merit
GCS	Ground Control Station
HCI	Human Computer Interface
HIL	Human-In-the-Loop
HOL	Human-Out-of-the-Loop
ISR	Intelligence, Surveillance and Reconnaissance
LOA	Levels of Automation
MAGE	Multiple-UAV Agency simulation testbed
MAS	Multi-Agent System
MC	Mission Commander (early def.)/Mission Coordinator (later def.)
NASA TLX	National Aeronautics and Space Administration Task Load Index
NDM	Naturalistic Decision Making
OODA	Observation, Orientation, Decision, and Action
ROV	Remotely Operated Vehicle
RPD	Recognition-Primed Decision
SA	Situation(al) Awareness
SAD	Search and Destroy
SAM	Surface-to-Air Missile
SAR	Search and Rescue
SEAD	Suppression of Enemy Air Defenses
SME	Subject Matter Expert
SO	Sensor Operator
UCAV	Uninhabited Combat Aerial Vehicle (subset of UAV)
UAV	Uninhabited Aerial Vehicle

1 Introduction

Although many attempts have been made to create a practical framework for the design of user interfaces, there has not been a definitive, easy to follow set of steps in design. Rather, the designer is told to understand the key concepts and then figure out how to translate that into a display.

Designing effective human-system interaction is critical when there are increases in automation and computerization in complex systems. The user is separated from the entity and can feel disassociated from the fate of the automated system. The increased automation can lead to a loss of situational awareness, because the operator may not feel entirely immersed.

Purely technology-centered approaches that consider the human as an afterthought may lead to user/system mismatches resulting in lack of acceptance, or worse – fatalities. There is an abundance of examples for this, such as the Three Mile Island meltdown attributed to pertinent information being buried down 8 menu levels; NASA landing system requirements calculated in meters while all other systems were in foot-pound measurements; and the engineering designs that never caught on because they were built for the engineer rather than the user, such as Reverse Polish Notation calculators. Clearly, design impacts usage. Technological advances must be integrated with human capabilities in the context of the application. This is one of the major tenets of Klein's (1993) Naturalistic Decision Making.

Our research framework is discussed in-depth in Chapter 4 and follows the McNeese, Bautsch *et al*. (1999) recommended research framework. It is an iterative approach that allows for an analysis of a wide range of complex systems research to proceed in a reasonably straightforward way.

This research's goal is to create a framework for display designers to systematically create cognitively sensitive displays – defined as a system in which the human operator can recognize a solution pattern without tremendous mental gymnastics, i.e., a high mental workload. Furthermore, this research presents one possible set of tools and a recommended ordering for applying those tools. The design toolkit includes Naturalistic Decision Making, Semiotic Analysis, Operator Function Model and Situation Awareness tools. With this toolkit, the designer should be able to consistently create displays of high quality and usefulness.

ASSESS
the display
design success

WHAT
information
needs to be
presented?

HOW
is the
information to
be presented?
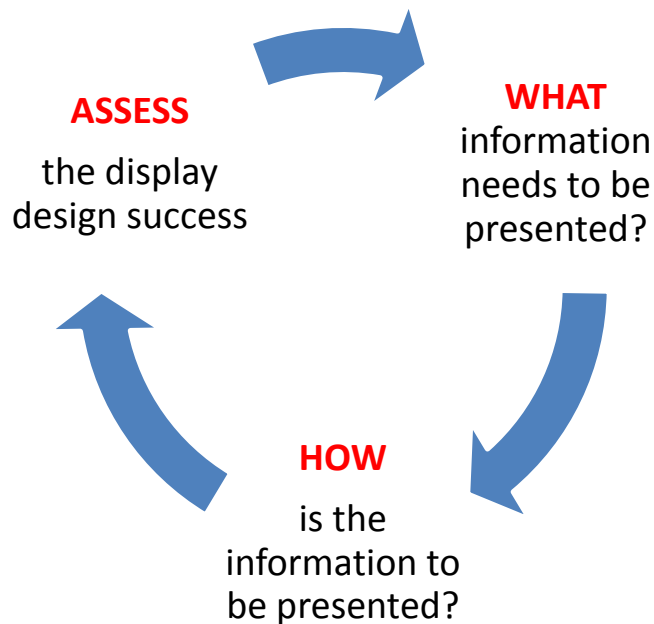
Figure 1.1 Iterative process to designing a display

The process for designing a display should be a simple iterative process that asks "What information needs to be presented?"; How should that information be present to the user?"; and "How does one evaluate the goodness of the design?" (Figure 1.3). Each time through the iteration allows a more complete tuning of the display into the final desired product. Chapter 2 will expand the discussion of this iterative process and where the tools in the toolkit can be applied. The toolkit has the flexibility of modular tools that can be changed to fit the problem and designer's knowledge. For example, if the designer is more comfortable with another methodology for eliciting information form stakeholders, then he can substitute that tool for the NDM toolkit.

1.1 *Research Model*

To perform any research in the Human Computer Interface (HCI) domain, the researcher must ask three very important questions. The first two questions are about the "what" and "how" of the information to present when designing the visual user interface, while the third question is interested in deciding the effectiveness or "goodness" of the HCI design. Another way of looking at the display design is the semantics (meaning) and syntax (form) of the message being sent by the UAVs to the operator/supervisor. We first elicit the appropriate or desired information for this domain (semantics) by interviewing subject matter experts in the field of UAV command and control; and then present it in a clear, concise way (syntax) through Visual Thinking design principles.

So the theoretical building blocks needed to create a cognitively sensitive display are "what", "how" and "assess", but what tools can perform these functions (Figure 1.2). This research used NDM to elicit the "what" from the SMEs and then used Semiotic

Analysis collaboratively to present some "back of the envelope" design ideas – bar graphs, gauges, etc. Further Semiotic Analysis applied with Visual Thinking visual cues (color coding, area visual acuity, etc.) suggested the first draft of a Visual Thinking sprocket – our "how". Finally, using SA techniques to compose questions, we showed the resulting images to non-experience subjects. Furthermore, the results were so favorable that the displays were demonstrated to active duty military SMEs for their comments.

Naturalistic Decision Making	Semiotics	Semiotics	Visual Thinking	Naturalistic Decision Making	Situational Awareness

What (Semantics)	How (Syntax)	Assessment

Cognitively Sensitive Display
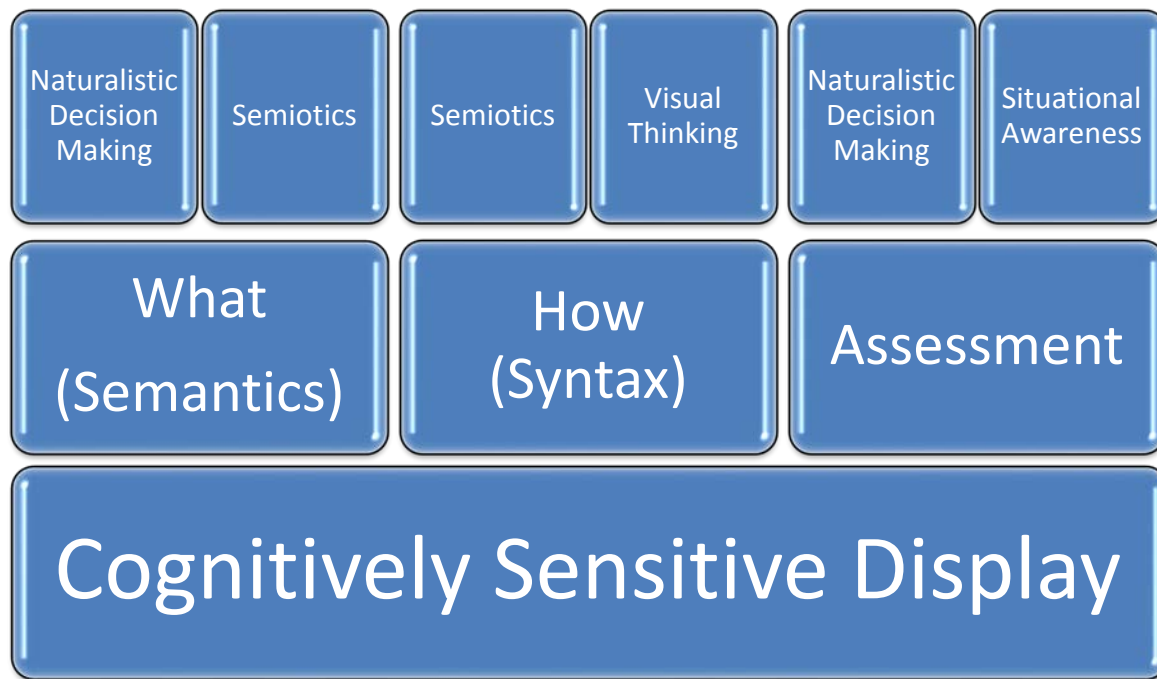
Figure 1.2 Experimental design building blocks for a cognitively sensitive user interface

The first question, "What information needs to be presented to the user?" examines the implications and content of the information pipeline, and makes tradeoffs. This solution incorporates the Naturalistic Decision Making (NDM) principle of developing designs with ample consideration to the domain users, i.e., develop products within a given context, not in a vacuum. Thus, subject matter experts (SMEs) were

interviewed to elicit what information should be included in any viable display, and these categories were then subjectively weighted by the SMEs. After interviewing and observing UAV operators and instructors at the Victorville, CA UAV flight school, a flexible, intuitive display was envisioned – all SMEs were either active duty or retired USAF UAV operators and the instructors represented the manufacturers. At least three levels of display flexibility were requested: USAF, manufacture and operator. In other words, the USAF may require certain Fields of Merits (FoMs), the manufacturer may require UAV model specific FoMs, while the operator may know operational/mission specific information that needs to be displayed. That is, the USAF might require the display supply the air speed for all missions, the manufacturer may require crosswind calculations to be displayed for this particular model, and the operator may want the oil temperature for this particular UAV because "it ran hot last mission". The design of the Visual Sprocket is such that the input stream can be changed to reflect "best practice" mandates directed by the organizations and manufacturers, while at the same time allowing for individual operator preferences. This design allows the flexibility the operator desires, while addressing organizational and manufacturer concerns.

For a visual display within the UAV mission planning domain, the pertinent information must be quickly interpretable and assist in narrowing down the possible alternate choices without excessively taxing the operator's mental capacity. At the same time, the underlying details behind the information must be accessible. Within this structure, the minimum / maximum / current / alert data values should be readily available to the decision maker. Considering that any decision will have multiple dimensions to consider and requires a quick response, then waiting for the operator to

"interpret" or "calculate" the raw data values on the display can be detrimental to resolving the decision.

The second question, "How is the information to be presented to the user?" examines the graphical layout and visual cues to be implemented or constructed. Visual Thinking posits that people make decisions almost instantaneously using visual cues in the environment. Through evolution, the eye/brain connection has optimized responses to certain visual cues (such as movement, shape, size, etc.). If the display follows Visual Thinking guidelines, then the display should require a minimal mental workload and able to be constructed with culturally obvious visual cues. The information sought by the operator should nearly "pop off" the screen, i.e., allowing a near gestalt interpretation of the information presented by the display. Semiosis (study of signs) supplies the relevant cognitive model for display cues, in particular, designing culturally relevant signs. The cultural influence of the signs this research explored were within the existing military culture for UAV ground control stations and international representations of cold and hot (water temperature) conventions. At the same time, it should be recognized that the user interface is the communication pipeline from the UAV to the operator (as envisioned by the UI designer).

Within the third question, "How do we assess the HCI?" is the issue of quantifying the user interface experience. Is the proposed display truly better than existing displays? NDM supplied the structure and tools to observe the single UAV operators "in the wild", and from those observations led to a new hypothetical multi-UAV display design. NDM also supports testing the final display results with Subject

Matter Experts (SMEs), while at the same time acknowledging the usefulness of non-domain subjects for controlled testing environments. Furthermore, situation awareness research has supplied the framework upon which this very subjective "is this display better" can be objectively examined. Designing the experiments to address specific levels of situation awareness allows for the quantification of the user experience. Finally, in accordance with the European Organization for the Safety of Air Navigation, Situation Present Assessment Method (SPAM) was used as it is one of the best experiment assessment feedback tools to use when addressing complex dynamic systems in this domain.

1.2 *Domain*

One of the goals of any strategic military commander is to inflict the most damage on the enemy with the least cost while sustaining the least amount of damage to their own troops. The easiest technological solution to this problem was to "extend the warrior's arm", meaning extending the offensive reach while at the same time not allowing the enemy to attack. For example, swords are preferred to daggers, spears are preferred to swords, archery preferred to hand-to-hand, etc. Each of these technological innovations placed the warrior farther from the opponent.

Remotely operated vehicles (ROVs) remove the human from harm's way, and are thus a technological innovation that enables this extension objective. Uninhabited aerial vehicles (UAVs) belong to the class of ROVs that function without a human crew on board. When the UAVs carry munitions, they are called uninhabited combat aerial

vehicles (UCAVs). Even though the human is not on board, humans play an important role in the applications of UAVs in practice (Narayanan, Edala et al. 1999).

Controlling a single UAV is a complex task: although there are multiple sensor inputs, many of those inputs are very limited in their ability to transmit data about the entire situation. As an example, consider the Predator UAV system (or squadron): "the fully operational system consists of four air vehicles (with sensors), a ground control station (GCS), a Predator primary satellite link communication suite and 55 people" (Wikipedia 2007). For each of the four vehicles, the three most important personnel are the pilot operator and the two sensor operators. Data supplied to the operators comes from the UAV's sensors and from external sources. Former USAF Deputy Chief of Staff for ISR, Lt. Gen. David Deptula, recently observed that a single Predator soon will be able to provide 60 or more video streams and noted that, "In terms of information[1] fusion, today's main challenge is not too little but rather too much information."

Furthermore, USAF Chief of Staff, Gen. Norton A. Schwartz, said that approximately 160 persons are required to perform "command and control" (C^2) and "processing, exploitation, and dissemination" (PED) for a single Predator. PED includes, but is not limited to, data interpretation, intelligence, mission planning, etc. Now imagine

---

[1] Throughout this dissertation, the use of the word "information" has the strict meaning of data plus meaning. In other words, the display shows the operator data that has been filtered through a semantic filter, placing that information with context (maximum, minimum, tolerances, etc.) The displays are not raw data that need further interpretation skills by the operator at the basic level. The operator may have to mentally join or split displayed information to assist current decision actions.

the added system complexity when multiple UAVs are "flown" by a single person – now in a supervisory role.

Obviously, if the operator controller is going to graduate to a supervisory role, then the information needed by the supervisor is significantly different. More of the rudimentary flying is offloaded to the UAV through automation, allowing the supervisor to make more of the difficult, mission decisions. This fits with the 2010 USAF Technology Horizons study's top finding, which identified the greater use of autonomous systems as critical to realizing capability increases. Within this emerging multi-UAV domain, developing new displays to present the information needed by the supervisor is a priority.

This research assumes the operator works in a supervisory role and interacts with the multiple UAVs in a method similar to a simulation game player. Directions are given to the entities – individually or as a group. The entities are semi-autonomous and perform tasks without interference from the supervisor – unless the supervisor needs to re-task the entities or the task requires human intervention (bombing, surveillance imagery, etc.). An entity may be re-tasked (externally by the supervisor or internally by the entity) because of successful completion of a subgoal, deficiencies in performance, or changed mission priorities.

Yet another question to ask is "Why do we need a new user interface?" Two simple answers to this question: (1) The current display for one UAV is barely sufficient to successfully operate the current control system of single UAVs – crashes during takeoff and landing happen much too frequently; and (2) The large amount of data

inundation / information overload for the UAV operator must be brought under control, especially when increasing the number of UAVs the operator supervises. Controlling multiple UAVs dictates that the operator-level "piloting" skills used with one UAV, such as takeoff and landing, must be automated and offloaded to the UAV. This automation has been incorporated in the aircraft of the domestic and international airlines for years.

The USAF clearly envisions autonomy as a major enabler of its vision for the future of RPAs. More importantly, a key assumption in the USAF UAS Flight Plan is that "Automation with a clear and effective user interface are the keys to increasing effects while potentially reducing cost, forward footprint, and risk" (HQ USAF, 2009).

1.3 *Research Approach*

This research developed a cognitively sensitive approach that efficiently incorporated the human into supervisory roles for mission planning and system monitoring for multiple UAVs. A cognitively sensitive system is defined as a system in which the human operator can recognize a solution pattern without tremendous mental gymnastics, i.e., a high mental workload. Another way of saying this is the solution is recognized as a whole, gestalt pattern, or "status at a glance". If the pattern does not exactly fit the requirements of the solution, then the operator can mentally reorganize the given pattern with relative ease to suggest an alternative solution. It is hypothesized that a system designed and implemented using this approach will improve operator performance especially in high-stress or time intensive situations.

This display system is contrasted with existing user interfaces within the military UAV domain, mainly tables of numbers, and what can be considered traditional graphical user interface design packages including analog gauges or bar charts.

In order to evaluate the displays, a high-fidelity test bed that could be configured for the alternate displays is necessary. This research utilized Sytronics, Inc. Multiple-UAV Agency (MAGE®) test bed and it was designed to integrate net centric information to facilitate control of multiple UAVs. One of the chosen features for the MAGE® software was to provide automated mission planning to reduce operator workload. Towards this end, Sytronics, Inc. licensed from Operations Research Concepts Applied, Inc. (ORCA) its OPUS® mission planning software library. OPUS® was employed by MAGE® to generate multiple alternative mission routes from which the MAGE® operator could choose to execute. In addition to the routes, the OPUS® software also generates a series of Figures of Merit (FoM) to describe the characteristics of each alternative. These FoMs could include more than a dozen dimensions, including the probability of surviving the mission, number of surface-to-air missile (SAM) launches, minutes of exposure to anti-aircraft artillery, minutes exposure to search radars, minutes exposure to missile guidance radars, fuel consumption, to name a few. These measures are all numeric but have different measures, different minimums and maximums, and run in different directions (fewer SAM missile launches and higher probability of survival are both better).

MAGE® uses state of the art agent programming to assist the operator, but the simulation has been limited to a traditional Microsoft Windows® / Java-style interface,

i.e., buttons, trees, menus, etc. This clunky user interface paradigm may be quite adequate for an office environment, but can be a hindrance to a time sensitive, high stress military mission. Requiring a user to traverse a menu system multiple layers deep or read several pages of data tables has been shown to be a human factors nightmare.[2] The research examined the user interface as a presenter of information in a format that caused little cognitive dissonance. Key FoMs of the problem to be solved by the operator are presented as a visual pattern that he/she can readily recognize and allowed him/her to quickly identify a potential solution.
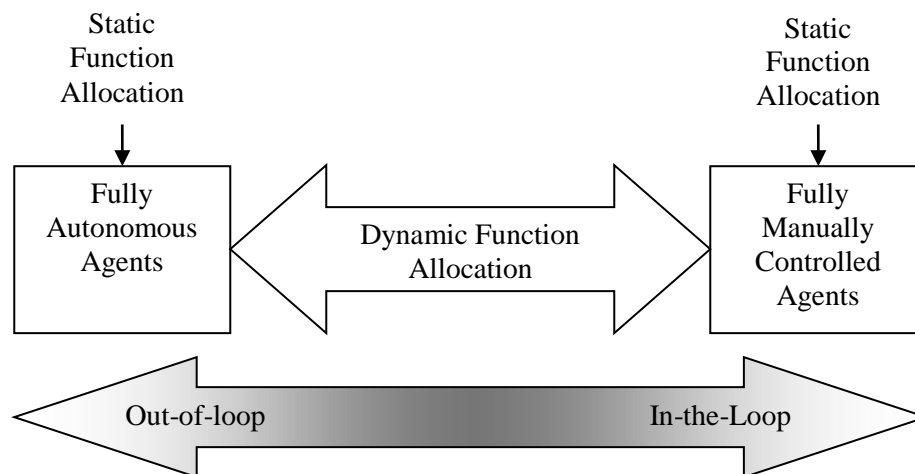
Figure 1.3 Research into the human's role in control

[2] *"On March 28, 1979 a sequence of events; equipment malfunction, design related problems and worker errors, led to a partial meltdown of Unit 2 at the Three Mile Island nuclear power plant near Middletown, Pennsylvania. The main feed water pumps stopped running, which prevented the steam generators from removing heat from the reactor.* **Signals available to the operators failed to show exactly what had happened***, which may be why they took a series of actions that made conditions worse by simply reducing the flow of coolant through the core."*

    Lloyd J. Dumas*, Waking up to the Real Threats to Security in the Post-Cold War World.* **Human Factor and the risk of Nuclear War** http://www.slmk.org/main/artiklar/Human_Factor.pdf, Svenska Läkare mot Kärnvapen (*Swedish Section of International Physicians for the Prevention of Nuclear War)* (Accessed March 15, 2008).

How does one human operator/supervisor control multiple UAVs? What control mechanisms or paradigms are needed to assure the appropriate level of human control? In other words, must the control mechanisms for controlling multiple UAVs be significantly different than the control mechanisms needed to control an individual UAV? Figure 1.3 illustrates current research at the ends of the spectrum with regards to controlling multiple agents. Those exploring swarm intelligence (Bonabeau and Théraulaz 2000; Dorigo, Bonabeau et al. 2000; Gaudiano, Shargel et al. 2003; Holland, Woods et al. 2005) assume a fully autonomous swarm without a human-in-the-loop to solve the given problem space. In their approach the human presents the scenario to solve to the swarm and the swarm proceeds to "discover" a viable solution. At the other research extreme (Narayanan, Ruff et al. 2000; Karim, Heinze et al. 2004; Ruff, Calhoun et al. 2004; Lewis, Wang et al. 2006), the researchers are not interested in swarm intelligence; but rather, are interested in humans controlling multiple autonomous agents. These researchers experimentally demonstrated that under various levels of autonomy, operators can control from one to thirteen UAVs. They claim that any larger group of UAVs would require a massive leap forward in UAV autonomy technology.

These laboratory results were discussed among software designers and Multiple-UAV Aircraft Control (MAC) Predator operators/pilots during the design of the MAGE simulation testbed. The MAC concept was proposed and deployed as a method to decrease the number of personnel needed to man multiple operational UAVs. The MAC was envisioned to be a group of four traditional GCSs, but manned by four sensor operators (SO) and one pilot – the pilot was the commander. As a method of reducing the pilot workload, the SO controlled the UAV during low-intensity legs of the mission.

However, after implementing the MAC, the pilot felt situationally detached during these low intensity operations. When the pilot had to control a UAV in a (planned) high-intensity leg of the mission (surveillance of a moving target or nearing a weapons release), an auxiliary pilot was required to attend to the remaining (low-intensity) UAVs. When a UAV unexpectedly goes from low- to high-intensity activity, the pilot felt he needed to be in the loop at least one-half hour earlier. Although a MAC was originally envisioned to have a capability of four UAVs being simultaneously controlled, only three UAVs were put into operation because of the pilots' fear of being overloaded. Finally, the geographic separation of the operator to the UAV (operator in Nevada and the UAV in Afghanistan) seems to aggravate many situation awareness problems. After experimenting and redeploying personnel in the 4-UAV MAC, the final empirical personnel ratio for a MAC was four SOs and three pilots – a savings of only one pilot!

When one discusses human factors engineering and user interfaces in complex, dynamic systems, we are entering the domain of situation awareness. Here we come to the essential/primary problem this research addresses – how to get the critical information to an operator in a timely manner. An operator has situation awareness of the system when he/she has reasonable knowledge of all critical aspects of a controlled system, i.e., informally, he/she knows what is happening. Endsley (1988) formally defined situation awareness as the "perception of the elements in the environment within a volume of time

and space, the comprehension of their meaning, and projection of their status in the near future"[3].

This dissertation is arranged in the following order: Chapter 2 is a review of the literature with respect to the military UAV domain, key technologies and the theories on which this research is based. The third chapter explores semiosis and how it was used to guide the research. The fourth chapter describes the research framework used throughout the research. The fifth chapter describes the research methodology for each experiment. The sixth chapter presents the findings of the research. And finally, the seventh chapter describes the research contributions.

---

[3] Notice the lack of reference to past actions and states. Gary Klein (NDM fame) told me at the 15th International Symposium on Aviation Psychology held at Wright State University that this alone is why his NDM is superior to Endsley's SA theory. The current situation must not be concerned with only the possible future outcomes, but rather reflection on past actions helps put the current situation into context.

2 Literature Review

This literature review is divided into two sections: (1) Theoretical Background associated with the design of a cognitively sensitive display and (2) Information about the multi-UAV domain necessary to understand the experiments – essentially, a toolkit and an object on which to use the tools.

Some of the past cognitive science theories that have explored the design of the user interface: Naturalistic Decision Making (Klein 1993), Visual Thinking (Arnheim 1969), and Situation Awareness (Endsley 2006). Each of these approaches tries to address the operator's strengths and weaknesses, each with its own level of success. Although the operator may have a goal or goals recognized by each of these theories, there is the distinct possibility that the operator goals may not fit the specific theory – the proverbial "square peg in a round hole". Additionally, to fill in some of the perceived applied design gaps, Semiotic Analysis[4] and Operator Function Models (OFM) were applied.

Klein theorizes that experts internally encode information based on experience; this experience is the experts' knowledge and that sets them apart from non-experts – the ability to "see/feel/hear" patterns in an evolving situation. Klein's research shows that experts do not necessarily compare multiple options, but rather use experience to look for

---

[4] Semiotic Analysis (semiosis) is the study of signs and their meanings. The display is a grouping of lines and symbols that are used to communicate between the system and the operator – as the designer understood the problem.

a solution (that has worked in the past) that satisfies the pattern of the problem without real consideration for optimal solutions. Arnheim claims all cognition is derived from visual processing and, because of this, displays should leverage this by encoding information consistent with visual perception capabilities[5]. He supports visual encoding of numerical information where applicable – maps that emphasize population density using color-coding, train routes using color and line width to emphasize direction and capacity, etc. Endsley postulates the existence of three levels of situation awareness and any interface that is going to succeed in any dynamically complex domain must address these three levels – unsuccessful interfaces will fail in at least one level of SA when presenting the operator with the needed information to make a correct decision.

These theories direct the designer to encode vital information the operator can easily recognize, that will in turn, allow the operator to make correct and efficient decisions. However, none of these three theories had a specific methodology for encoding the display information. This is obviously a distinct shortfall for these methods to the practical designer, but later authors have addressed some of these failings. In particular, the design of the display utilized semiotic analysis (Nadin 1988; Tufte 1997; Ferreira, Barr et al. 2005; Tufte 2006) and recommendations from information / scientific visualization to examine what might be considered culturally significant visual cues (Tufte 1997; Spence 2001; Ware 2004; Tufte 2006).

---

[5] This seems a little extreme, since by this definition, blind people could not reason! I believe it is meant that most survival and base instincts/decisions are predominately visually aided, i.e., fight or flee decisions. In this context, slow thoughtful decisions can have fatal consequences.

The goal of this research is to create a framework for designers to systematically design displays. The iterative steps to creating the display are as follows:

1) Interview stakeholders within the domain to ascertain what information needs to be presented [Naturalistic Decision Making (NDM) and Semiotic Analysis];

2) Develop a user model [Operator Function Model (OFM)];

3) Use semiotic analysis to create a display that encodes the information with visual cues. These cues should be design such that the human can quickly recognize any patterns [Semiotic Analysis and Visual Thinking (VT)];

4) Evaluate the design empirically through laboratory experiments [Situation Awareness (SA)] and "in the wild" with stakeholders [NDM]; and

5) Assess the success of the designed display with stakeholders [NDM, SA and VT]. Repeat if necessary.

An instantiation of the objective of this research was to develop a cognitively sensitive approach that incorporates the human into the decision mechanism of supervising multiple UAVs to promote effective control. In the context of human centered decision-making of multiple UAV control, this research explores whether a "cognitively sensitive" paradigm can be utilized to design intelligent user interfaces that are human friendly. It is hypothesized that a system developed using this approach will improve operator performance in high-stress situations.

2.1 *Theoretical Background*

The research is presented in the order necessary to successfully proceed through the steps of the framework proposed in Figure 1.3. The steps are expanded in Figure 2.1. As can be seen by Figure 2.1, the three steps have been expanded to five steps, but in

reality, the "what" and "assess" steps are expanded to show the tools used, and is not necessarily sequential. In fact, in this research, the assessment of the results was done using an in-house pilot study with a reassessment of the display before the formal experiments were performed. This pilot study inspired a mouse rollover function to display the underlying data – minimum, maximum, low and high thresholds, etc. This rollover function was not tested as part of the experiment, but was added to subsequent releases at the behest of other stakeholders. The steps for designing a display should lead to an iterative solution.

ASSESS
Assess success of the designed display w/ stakeholders [NDM]. Repeat if necessary.

WHAT
Interview domain stakeholders [NDM] for what information needs to be presented [Semiosis]

WHAT
Develop an operator function model (OFM)

ASSESS
Evaluate the design empirically: laboratory [SA] and "in the wild" [NDM]

HOW
Use semiotic analysis to create a display that encodes the data w/ visual thinking [VT]

Figure 2.1 Steps to designing a cognitively sensitive display

Furthermore, the three steps for designing are not temporally mutually exclusive – they frequently overlap. However, this suggested guide can keep the designer on task enough to say that they are currently performing step X.

The organization of Section 2.1, as mentioned previously, is illustrated in Figure 2.1. Section 2.1.1 discusses Naturalistic Decision Making (NDM) and how it can be used as a tool to extract **What** information needs to be presented to the user in the new display through interviews and examining the existing context or environment. Furthermore, NDM can help validate the display, and verify the user has the information needed. Section 2.1.2 presents a user modeling tool, called the Operator Function Model, which will be used in Chapter 4 to more completely understand what the operator needs to know so that they can make quick and accurate decisions. Semiotic analysis is described in Section 2.1.3 and used as a tool in Chapter 3 to perform an analysis of the existing and proposed alternate displays. The analysis helps transition from the **What** to the **How** steps of the process. Following the analysis, the research examines Visual Thinking (Section 2.1.4) and its implications in the **How** information is presented to the user, addressing the need to present information that the human can perceive and categorize quickly. The last step of this process is to **Assess** the display within the context of the domain. To achieve this, we examine the Situation Awareness research as a measurement tool (Section 2.1.5). Furthermore, NDM is again used to verify the display is useful within the context of the domain. Finally, Section 2.1.6 examines the research on Separate, Integral and Integrated Displays and how these concepts can help explain the research decision process on which existing and proposed alternate displays were chosen and why.

2.1.1 Naturalistic Decision Making

In the first step of the cognitively sensitive display design, one needs to gather information about "what" should be presented to the operator/supervisor. This is one of the most important steps, because an incomplete picture of the problem will probably produce an incomplete solution. One such tool that has been successfully used within the human factors community is Naturalistic Decision Making (NDM).

In his report on **Naturalistic Decision Making**, Klein (1993) asserts that most designers of decision aids are handicapped by a lack of knowledge of the decision maker's internal decision algorithms for any given domain. Human computer interfaces (HCIs) being developed "to help people perform cognitive tasks do not support decision making". Just presenting the information is not enough. It has to be presented in a cognitively sensitive way.

The expert decision maker may not have all the information he would like, but because of constraints (time, cost, etc.) a decision still must be made. The expert infers or gathers information from external sources, "rules of thumb", experience, etc. As described by Klein (1993), experienced people do not normally perform a formal decision analysis, but rather, attempts to pattern match previous situations to identify potential solutions. The solutions are adapted to fit the current problem.

On the other hand, the designer has formal specifications that present the apparent minimum decision requirements needed by a neophyte – requirements that do not necessarily provide the whole cognitive picture. From these specifications, the designer

is expected to design an HCI without being given information about how the operators expect to use it to make their decisions. The resulting decision aid may not have the "feel of the system" that an expert learns over time.

NDM attempts to go beyond the formal specifications by exploring how (and why) operators make important decisions. It looks at the decision requirements of an expert, not a neophyte, and then presents that information to the operator. One such tool is Recognition-Primed Decision-Making (RPD).

2.1.1.1 Recognition-Primed Decision Making

One naturally assumes that when a person makes a complicated decision, he would mentally make a list of alternative solutions and then narrow it down to two or three good choices. Then these small few choices would be examined in more depth to select the best.

Klein (1993) found that this is just not the case. Instead, he found that experienced problem solvers looked at the current problem and tried to find a previous problem that was similar. They then mentally tried to use the successful solution on new problem – if it did not work mentally, then they tried to adapt it mentally to better fit the new problem. In other words, they recognized patterns in the problem to find a solution template in the solution space. New innovative solutions were not created from scratch. Alternative solution lists were not created.

These recognized patterns are the priming needed to find a viable solution. It does not necessarily guarantee an optimal solution, but rather a tried and true solution

that satisfies the problem constrains. Keeping the patterns simple for the operator / supervisor allows them to quickly assimilate the situation – patterns such as "bigger is better" and "best is a medium circle shape".

2.1.2 User Modeling

Because the first tool does not have a formal method of modeling the expert user, a second tool is needed to do this. NDM does recognize the need to understand the expert and their decisions, but it does not have a specific toolset to perform this modeling.

There exist several tools to that do such as Task Analysis (McCormick, 1976; Sanders and McCormick, 1987) and Operator Function Model. This research's primary user model is the Operator Function Model (OFM), with the proposed solution for this domain presented in chapter 4. Discrete control models (task analysis) and models of operator function using discrete control modeling (OFM; Mitchell and Miller, 1985, 1986; Mitchell, 1987, 1996) constructs have been successfully used for modeling users in complex systems. Task analysis identifies "and list[s] all the human operations performed and their relation to system tasks (McCormick 1976, p. 24)."

They describe operator behavior in a range of complex systems and prescribe operator functions by representing the interrelations between dynamic system state and operator functions, subfunctions, control actions, and information needs related to operator activities (Mitchell and Miller, 1986; Mitchell, 1987, 1996). The OFM is an alternative to task analysis techniques used by other human factors engineers

(McCormick, 1976; Sanders and McCormick, 1987), since OFM is a modeling tool that provides a dynamic, task-analytic structure that can be used by system designers to define a user interface that is based on operator rather than hardware function (Mitchell, 1987, 1996).

2.1.2.1   Cognitive Models

From nonscientific anecdotal accounts (Flying Magazine) and accident reports [(Giffen and Rockwell 1987) summarizes many accident report findings], one finds that there is a certain mysticism associated with regards to a pilot's decision making skills. These reports find that poor decision making skills cause or contribute to many of the "pilot error" accidents. Most of these accidents involve time constraints or time pressure requiring a very rapid decision by the pilot; the reports' general recommendations tend to be "more research is needed on pilot judgment and decision making".

Despite these recommendations, researchers have been slow in examining the cognitive processes of the cockpit crew. In the absence of experimental/empirical data to support a new analytical model of decision making, it has been assumed that standard analytical model fits. Besco, Maurino *et al*. (1994) claimed:

> Decision making in the cockpit follows traditional views of decision making … in which the decision maker, i.e., Captain, is 1) presented with a situation that requires a decision; 2) the nature of the situation is assessed by the decision maker; who 3) determines the availability of alternative outcomes to respond to the situation, and 4) after evaluating the risk and benefits of each alternative; 5) selects an alternative in response to the needs of the situation. (p. 43)

The assumption that decision-makers' always examine all alternatives exhaustively is unrealistic. The Captain typically does not jot down a few alternatives,

24

weigh and balance each possible outcome, and then select the "best" alternative. Research in NDM argues that the Captain, if he/she is a good decision maker, is going to recognize the current situation fits a previous situation problem pattern and use (adapt) the solution that satisfied the previous situation, if possible. This requires much less cognitive workload from the Captain.

2.1.3 Semiosis, Semiotic Analysis

Semiosis is the study of signs; including the cultural influences on the interpretation of signs (Table 2.1). As an example, within the USA culture, the color red is often interpreted as danger, or bad luck; while in China, red is considered very lucky. The number "7" is lucky and "13" is unlucky in the USA, while in China, "4" is unlucky because of the word "four" in Chinese sounds the very similar to the word "death". So a more complete definition of semiosis is the study of signs within a culture, but the culture does not have to be based on national or ethnic backgrounds. It can be the culture within a corporation (e.g., "mac happy face", "blue screen of death"), a military branch (iconology for war games), etc. In this section, the cultural influence of the signs this research explored were within the existing military culture for UAV ground control stations (GCS).[6]

---

[6] A detailed semiotic analysis of this display is reported in Chapter 3.

Table 2.1 Some definitions of semiotics and its uses.

Semiotics is the doctrine of *signs*. The sign is the most important building block to semiotic study and it is defined as anything that stands for something else to some interpreter.	(Ferreira, Barr et al. 2005)
The object of semiotics is sign systems and their functioning within culture.	(Nadin 1988)
The main tenet of Semiotic Engineering is that interactive systems designers actually communicate with users (at interaction time) through computer systems interfaces. Interfaces act as the designers' proxies (the *designers' deputy*, according to the theory). Thus, when designing any system's interface, designers are actually deciding what kinds of conversations they will have with users, using which modes and media, and for what purposes.	(Valente, Souza et al. 2008)
Semiotics, also called semiotic studies or semiology, is the study of sign processes (semiosis), or signification and communication, signs and symbols,	(Wikipedia contributors 2010)
The overall goal of the human operator modeling and semiosis effort is to decide upon the content and form of information to be displayed to a well-trained and motivated ... supervisor .... Semiotic analysis deals with the assessment of the syntax, semantics, and pragmatics associated with human/system interaction.	(Narayanan, Ruff et al. 2000)

The history of semiotics has two major figures which have defined the western traditions, mainly the Swiss linguist Ferdinand de Saussure (1857-1913) and the American scientist and philosopher Charles Sanders Pierce. This research used a model derived by Pierce that has been shown to be appropriate for computer based signs. This model has a three part relationship containing the *representamen*, the *object*, and the *interpretant* (Figure 2.2). The representamen is the physical instantiation of the symbol in reality – like an eight sided red and white stop sign is physical representation of the concept to stop. The object is the actual concept for which the representamen stands – "cars stop here". And finally, the interpretant is therefore the sign created in the mind of the perceiver, or how the sign is perceived. To an observer, the representamen creates in the mind of the observer an equivalent or more "developed" sign – "I should stop here".
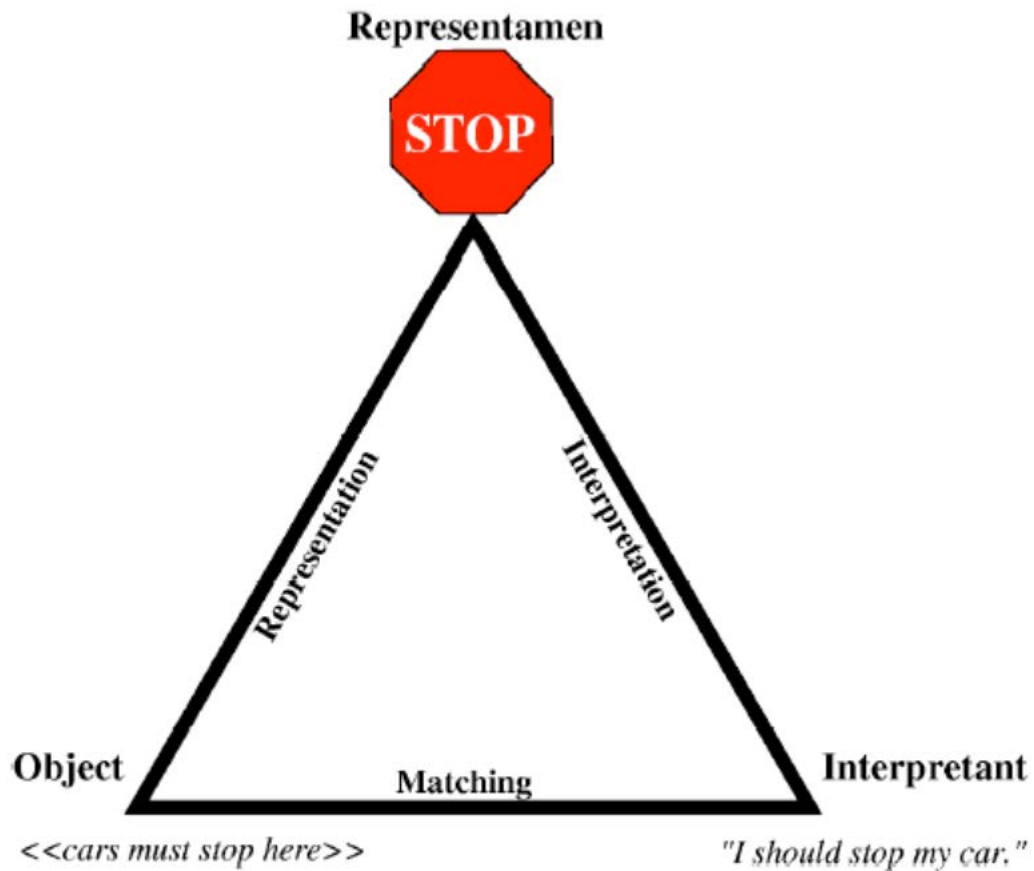
Figure 2.2 Peircean Triad

Although Peirce classified thousands of sign categories, he found the three most fundamental sign divisions are *icon*, *index* and *symbol*. Any given sign can be assigned one of these three categories based on the relationship between the object and the representamen (See Figure 2.3). These categories are not mutually exclusive; any given sign may belong to more than one of these categories.

In the left illustration of Figure 2.3, an iconic sign is presented. The representamen resembles a portrait of the author and the perceiver can interpret this as such precisely because the representamen looks like the author enough to be recognized.

It should be noted that the representamen can be (and has been) weakly argued to be an indexical sign precisely because it is not three dimensional.

Iconic Sign	Indexical Sign	Symbolic Sign

| Portrait of MJF | This is portrait of MJF | Weighs 1 kg | Weight shown on the scale is 1 kg | Do not do something | Something here is forbidden |
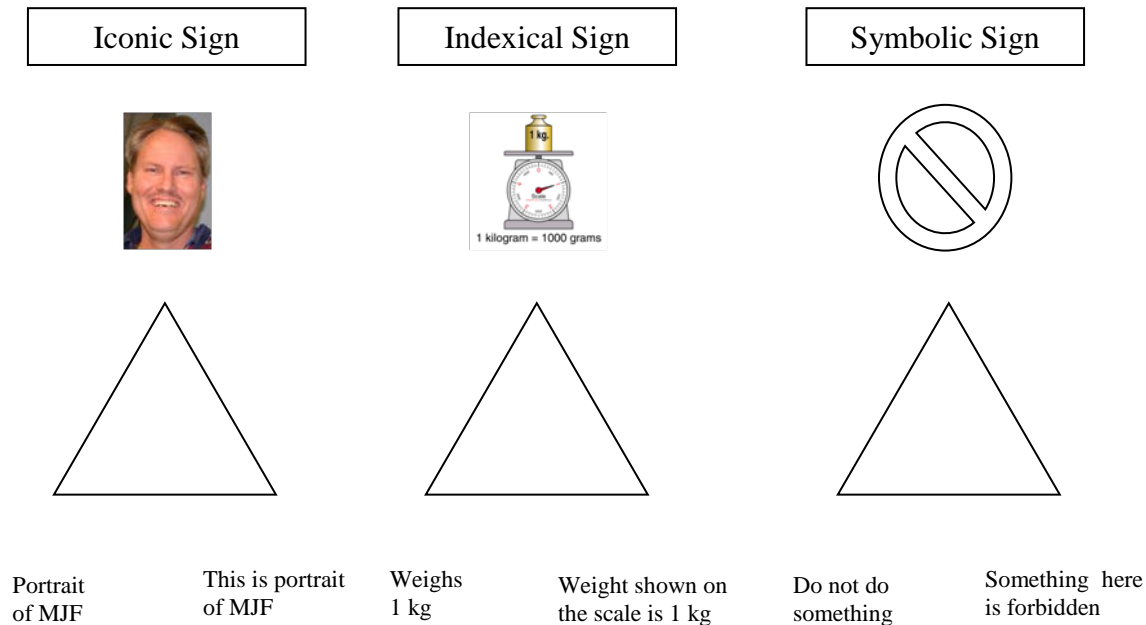
Figure 2.3 Sign classifications

The middle illustration shows an example of the indexical sign. In this case, the sign is representative of a (class of) object(s). It does not represent a specific object, but rather it makes a connection to a (class of) object(s) within the perceiver's mind. The object may not even be physically observable. In Figure 2.3, the middle illustration can be interpreted differently based on the context, the symbol could be interpreted as "weighing" (weigh something), "metric weighing" (weigh something using the metric system), "measurement" (measure something to exactly 1 kg, like in a recipe), or a class of measuring devices (scales).

Finally, the right illustration is an example of the iconic sign. These tend to be more culturally interpreted and are a learned interpretation. The sign on the right when

overlaid on another sign means "do not". For instance, placing this sign over an image of

a cigarette is interpreted as "do not smoke cigarettes near this sign", placing this sign over

the icons for restroom on a map should be interpreted as no restrooms available at this

site, placing the sign over a symbol of a fishing pole should be interpreted as "no

fishing", etc. However, without the interpretant being instructed by their culture as to the

meaning of the sign, there is no intuitive, obvious interpretation. With the training, the

interpretation of the overlaid sign is easy to generalize of other settings. It should be

noted that written languages are iconic – we are taught to read and write for many years

to understand this very complex signing language.

2.1.4 Visual Thinking

The Visual Thinking phrase was first coined by Arnheim (1969) to address how

people make decisions almost instantaneously. Inherent in vision is the ability to

preprocess data and recognize visual patterns.

> Vision is not perception and perception is not thinking. The mind
> gathers information and processes it. Note that I said information,
> information is data plus meaning. Before the mind conveys the
> information your eyes must observe it, and some preprocessing needs to
> be done to turn this data into information.

> Arnheim 2004

The key to Arnheim's thesis is that vision and thinking are not necessarily disjoint

concepts. When a person perceives an object with your eyes, before deep thoughts about

the object can be conceived, the simple sight of that thing at least causes classification

(placing the object in the context of other objects like it). For instance, if you see a cat,

before any separate thinking is performed about the cat, it has already been placed in the category of "cat". This is a particularly useful cognitive trait to have when that cat is a dangerous one that needs to be fled from, such as a tiger.

Arnheim contends the idea of visual thinking is an old one, going back to the ancient Greek philosophers: Plato, Socrates and Aristotle. These philosophers were the first to make a distinction between perceiving and reasoning, mainly because perception from direct senses could not be trusted. (We have all experienced "our eyes playing tricks on us", or heard tales of mirages in the desert.) Reasoning was considered to be the "correction of the senses" and the "establishment of truth".

It can be reasonably argued that Arnheim's visual thinking is almost instantaneous pattern classification. It is not the perception of the object that classifies the object, but rather the very well-travelled mental pathways that react with almost lightening quick classification. The perceiving of the object (cat) does not require new neurons to fire off and create new paths; the existing short pathways have always succeeded previously.

Within this same paradigm of visual thinking may fall the education concepts of audio and visual learning. It is recognized within the field of teaching that some students learn best by listening, others by seeing, and still others by doing. It is recommended that teachers make the effort to discover each student's learning style, and furthermore, present the subject material in as many modalities as possible to help the students' diverse learning styles.

2.1.4.1 Graphical Displays

This research proposes a new graphical display "widget" called the "visual thinking sprocket" to enhance the supervisor's control capabilities. To judge the appropriateness of the display, we look to Tufte's (2006) [p. 13] definition of what determines an excellent (statistical) graphical display:

- show the data
- induce the viewer to think about the substance rather than … something else
- avoid distorting what the data have to say
- present many numbers in a small space
- make large data sets coherent
- reveal the data at several levels of detail, from a broad overview to the fine structure
- serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- be closely integrated with the … data set.

The sprocket display fulfills all of these characteristics.

2.1.4.2 Visual Thinking Sprocket

Physiologically, the eyeball as an information-gathering instrument scans the world under the guidance of cognitive attention centers. The eyeball fixates on a region of interest. An image is buffered and scanned, like a massively parallel computer, to find objects within the image through feature extraction. Once extracted, these objects are serially scanned at about 25 items per second. Since the eye scans quickly, reacquiring a new image about 10 times a second, only four to twelve objects are recognized before the eye jumps to another fixation. These physical boundaries must drive the design of cognitively sensitive displays.

Furthermore, when designing a display, the two attributes must be balanced: the overview of the situation and the details within the situation. The overview is a qualitative "aspect of data preferably acquired rapidly and even better, pre-attentively; that is, without cognitive effort" (Spence 2007). A well designed overview display uses visual cues that are acknowledged to be pattern classifier aids so information "pops out" at the operator. On the other hand, details are quantitative and should only be presented to the operator on an as needed basis, i.e., when the operator requests more in-depth information, presumably because of the overview display observations.

Within the design of the Visual Thinking Sprocket display, primary attention is devoted to the overview pattern classifier aids. A design that presents an overview of a situation must be designed simply and stress those features that can be pre-attentively processed. According to Ware (2004), features that can be pre-attentively processed can be organized in the following categories:

- Form: Line orientation, line length, **line width**, line collinearity, **size**, **curvature**, **special grouping**, blur, added marks, numerosity
- Color: **Hue**, Intensity
- Motion: Flicker, **Direction of Motion**
- Spatial Position: 2D position, Stereoscopic depth, convex/concave shape from shading

The features in **bold** were the pre-attentive cues this Visual Thinking Sprocket research attempted to model.

With this in mind, let us examine the interesting history of the Visual Thinking Sprocket. One of the earliest applications of a sprocket design (Spence 2001) to graphically present data is Florence Nightingale's **Rose** (See Figure 2.4). During the

Crimea war, Nightingale visited the field hospitals while attending the sick and wounded. She was appalled at the squalid conditions at the hospitals and persuaded the Sanitary Commission to undertake improvements.

AT SCUTARI AND KULALI, FROM OCTᴿ 1ˢᵀ 1854. TO SEPTᴿ 30ᵀᴴ 1855.
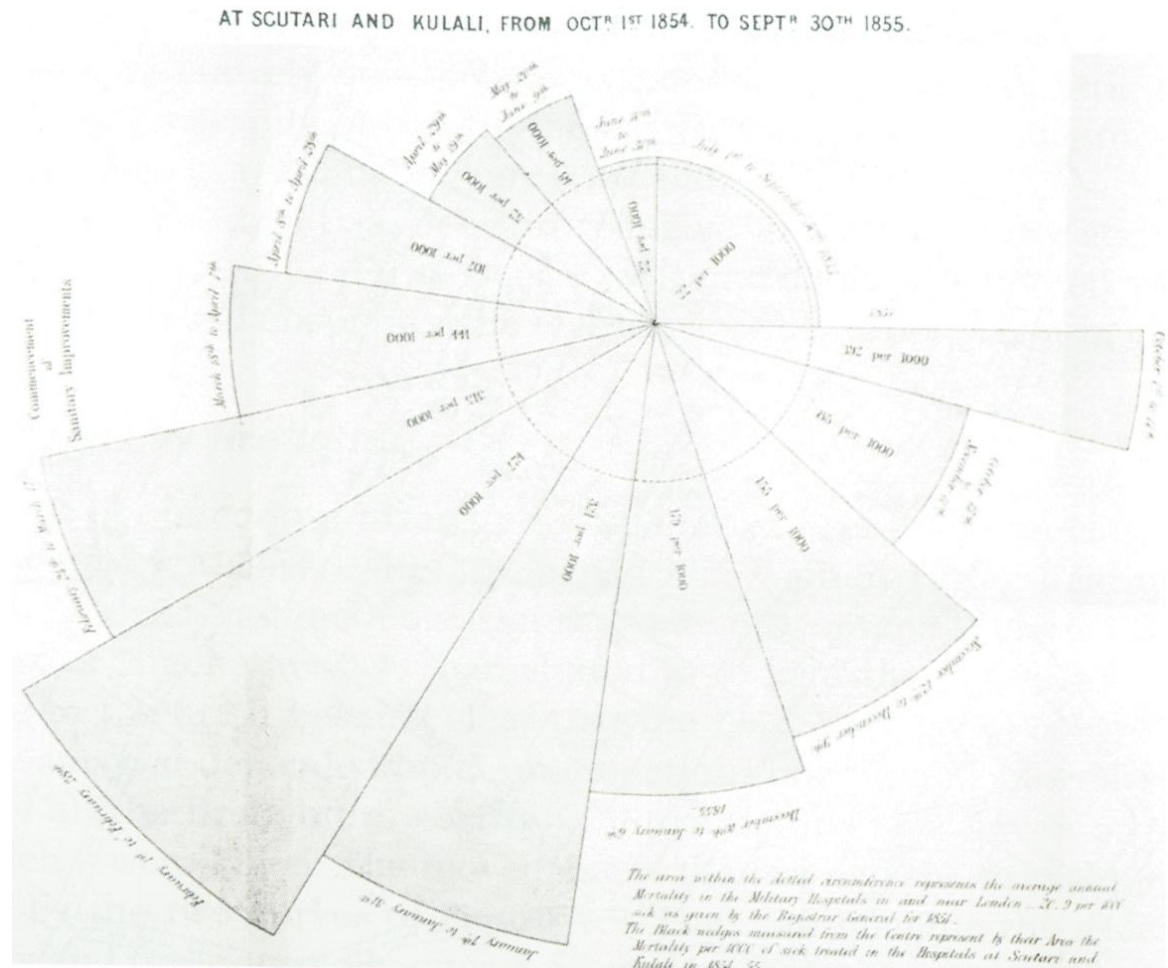
Figure 2.4 Nighingale's Rose illustrating the dramatic reduction of deaths in Crimea War field hospitals attributed to improving sanitary conditions

The Rose depicts the striking improvements attributed to the improved conditions: the length of each petal is proportional to the number of deaths that month; the subtended angle of each segment is the elapsed time; and the dotted line is the number of deaths in Army hospitals in Manchester, England. The new regime was initiated in March 1855

(approximately 9:00 o'clock on the Rose), and the figure dramatically shows the improvements.
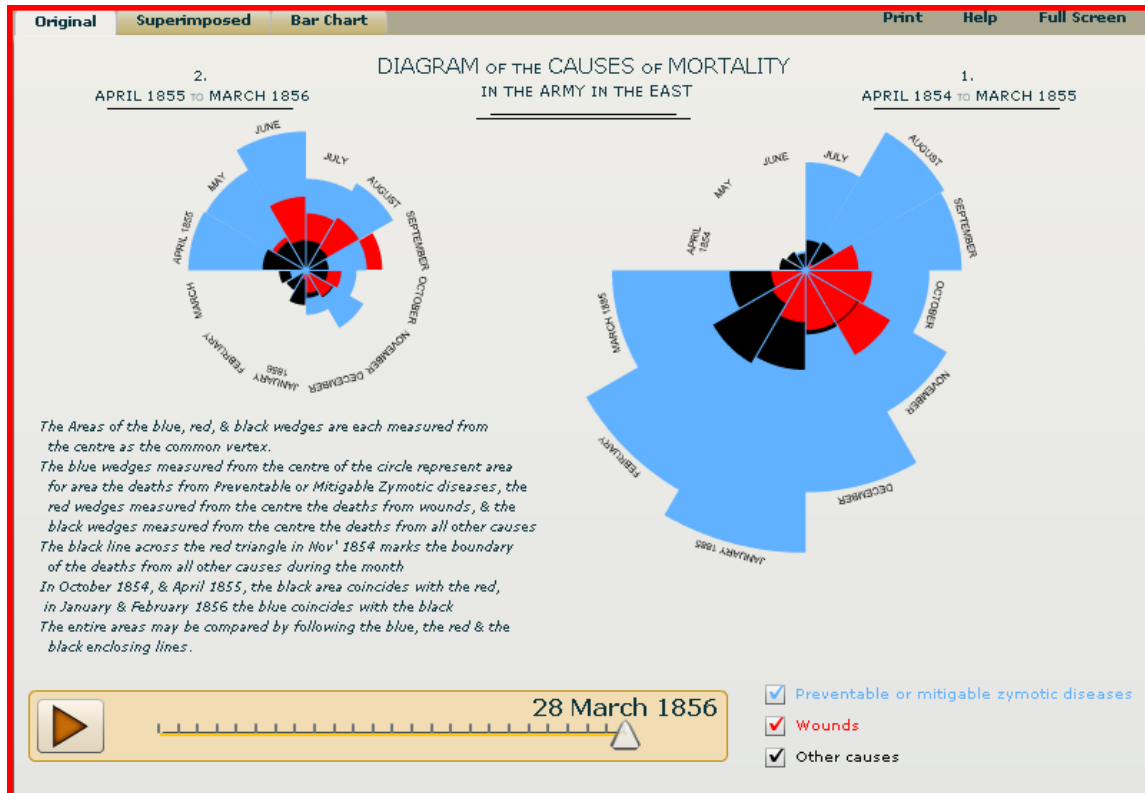
Figure 2.5 Interactive Nightingale's Rose (http://understandinguncertainty.org/coxcombs)

The Rose is such a fascinating visual tool that it has inspired several online versions of the display. One such interactive display is illustrated in Figure 2.5. With this display, the rose can grow throughout the Crimean War, each wedge being added, starting with April 1854 and proceeding through March 1856. The right-hand Rose precedes the left-hand Rose, with the first three months of the right-hand Rose being deaths preceding the Crimean War. The left-hand Rose is the second year of the war and illustrates how implementing the recommended sanitary guidelines greatly enhanced the chance of survival in the field hospitals.
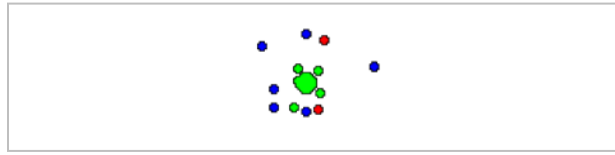
Figure 2.6 Starfield visual thinking widget

Singers and Endres (1996) present a second visual thinking widget they call the **Starfield®** (Johnson Controls Inc. 2001). The Starfield shown in Figure 2.6, was created to "help operators readily extract useful information from the vast quantities of data generated by complex systems", in particular, the facility management systems. A large facility can have dozens of zones, each with its own heating, ventilating, air conditioning, fire, lighting, and security systems. Text displays *can* provide values at each point in the entire system. However, an operator of a large facility may need to scan through "numerous floor plans or hundreds of thousands of lines of text" to find that one piece of relevant data. The Starfield represents data as points that appear in a scatterplot – "like stars in the sky". The clustering of points indicated data similarity or patterns of interest, with the color of the points indicating tolerance status [red (corresponding analog value is above the specified range), blue (below the specified range), green (within the specified range), gray (offline)]. Using a mouse to select a point provides more details about the data, for example doing a "mouse-over" of a point produces a Windows' "tool tip" indicating the name and value of the point while clicking on the point expands the details to include possible fixes. The largest dot represents the systems calculated global status. The most important parts of this research were: (1) operators are provided an "at-a-glance visual gestalt" of the data; (2) the display works to the humans strengths of pattern recognition and spatial reasoning; and (3) the display encourages data exploration.
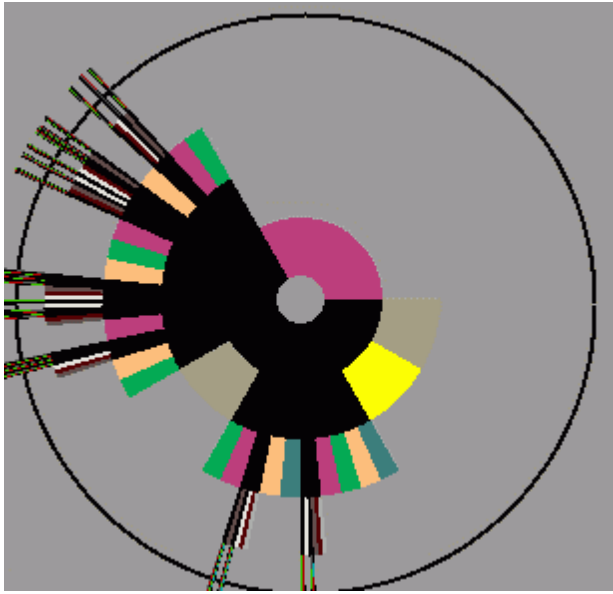
Figure 2.7 Visual thinking Kaleidoscope® for selecting search strategies

A third example of using a visual thinking widget, called a **Kaleidoscope** (Figure 2.7), is presented by Pu and Lalanne (2002). This Kaleidoscope is used as a decision tool to select land usage in a neighborhood. Existing structures (housing, cemetery, school, dumpsite, etc.) constrain the placement of new structures in available lots. For instance, do not place a dumpsite next to a school. In the Kaleidoscope, the entire circle represents the multivariate search space. Any solid lines represent a successfully met constraint, with longer (black) lines being better solutions. Extending through the edge are the potential solutions that must be examined in more depth.

Examining Figure 2.8, one can see the intentional feature implementation on the initial single-threshold Visual Thinking Sprocket design prototype. This Visual Thinking Sprocket was intended to be a decision support aid within a larger flight simulator.

36
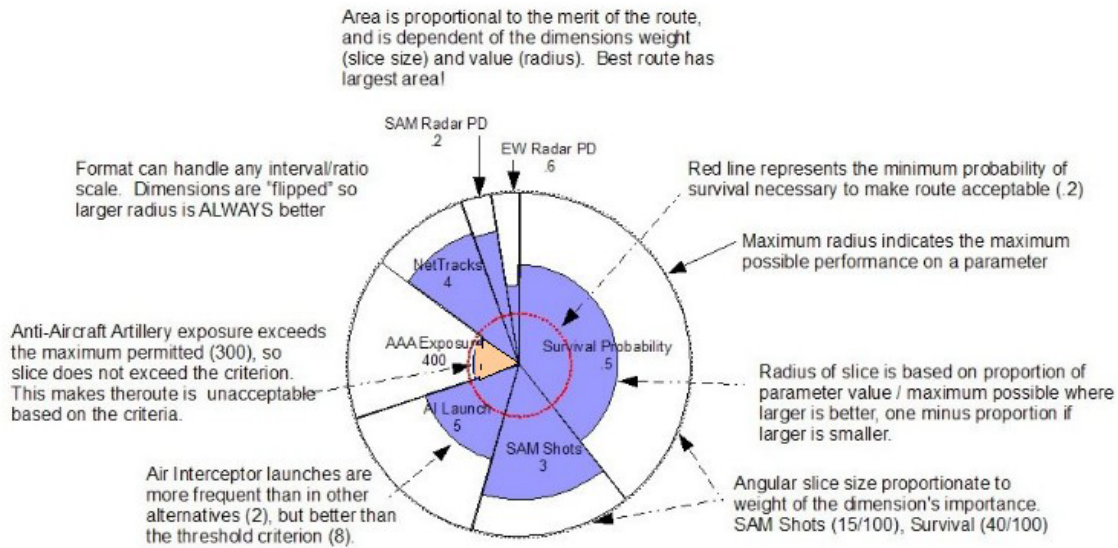
Figure 2.8 Early drawing of a multi-dimensional, multiple scaled decision support display. The raw detail data display is visible by mouse roll-over of the slice.

Encoded into this initial Visual Thinking Sprocket were (1) angular slices proportional to the weighting of the dimension; (2) acceptability of specific dimensions (pink – unacceptable, blue – acceptable); (3) individual dimension "health" or "preference" (larger colored area is always better); (4) slices nearer the red tolerance line are less optimal, those nearer the maximum radius are deemed near optimal; (5) labels naming individual dimensions and their associated current values; (6) a normalized rescaling of the dimensions; and (7) the global preference of the decision – bigger sprockets are better than smaller sprockets. Finally, if the operator wanted more information about a specific dimension, a simple "mouse-over" displays the detailed raw data behind the image.

From Figure 2.8, one can see why the resulting circular figure is called a sprocket, with geared teeth of varying length, resembling the tooth embellished wheel that drives a chain, or in this case, cognitive understanding.

2.1.4.2.1 Cognitive Congruence

Presenting information to a user in an easily decipherable context is difficult. As Tufte (2006) quoted from **The Commercial and Political Atlas** (London, 1786):

> Information, that is imperfectly acquired, is generally as imperfectly retained; and a man who has carefully investigated a printed table, finds, when done, that he has only a very faint and partial idea of what he has read; and that like a figure imprinted on sand, is soon totally erased and defaced... [pages 3-4].

This emphasizes the importance of imparting information in a fashion that is easily understood in an effort to make it effortlessly retainable. To transfer the information to the user, the user must grasp its meaning quickly without being overloaded with extraneous minutiae.

In educational parlance, people learn through three modalities: visual/spatial, auditory/sequential and kinesthetic/tactile (Silverman 2006). Visual learners prefer images, symbols, diagrams, etc. as the information conduit. Concepts are holistically understood. Thinking may be visualization three dimensions and occurs all at once rather than sequentially. Auditory learners remember much of what they hear (and even more about what they hear and then say). The auditory learner is a sequential thinker, preferring to "follow a logical pattern." [As an interesting aside, Felder and Silverman (1988) assert college age students tend to fall into the first category, while professors tend to fall into the second category.] The third modality, kinesthetic/tactile learn through touching or moving things. Learning is "anchored in the physical senses" and example or experimentation tends to help the learner. For example, infants must touch or put things

in their mouths to learn or understand them. Furthermore, athletes develop muscle memory by repetition of drills and exercises – this too is learning.

2.1.4.3 Concept of Cognitive Impedance

Cognitive impedance is the human information processing analog to electrical engineering concept of *impedance matching*. Cognitive matching refers to information presentation/representation coinciding with the recipient's internal model. If the representations are too discordant, the recipient must perform mental gymnastics to bring the representations more aligned. As examples, consider the electrical matching of two stereo systems and the problem of American aircraft horizon displays:

2.1.4.3.1 Example: Stereo systems in cars (4 Ohm) vs. home (8 Ohm)

To get the best sound from a stereo system, use components (amplifier, speakers, etc.) that have their Ohm ratings matched. In the American market, there exist 2 types of stereo systems: those made for automobiles are 4 ohm (4Ω) and those made for in-home are 8 ohms (8Ω). Matching the ohmage creates the least amount of audio distortion when the signal is sent from the amplifier to the speaker. If the components are mismatched, that is a 4Ω (8Ω) amplifier is paired with an 8Ω (4Ω) speakers, then "dropout" or distortion results in the speaker output. Granted, the $4\Omega => 8\Omega$ distortion (dropout) is different from the $8\Omega => 4\Omega$ distortion (overmatching, over amping, etc.), but they both will lead to signal distortion of different frequencies.

2.1.4.3.2 Example: Russian versus American Horizon displays

In the flight instrument that shows whether the airplane is flying level or not, the Russians use an outside-in display and the Americans use an inside-out display. What this means is that the perspective of the pilot relative to the aircraft is either from inside the plane to the horizon, or outside the plane from the horizon (Figure 2.9).

A) Inside-Out depiction of roll   B) Outside-In depiction of roll
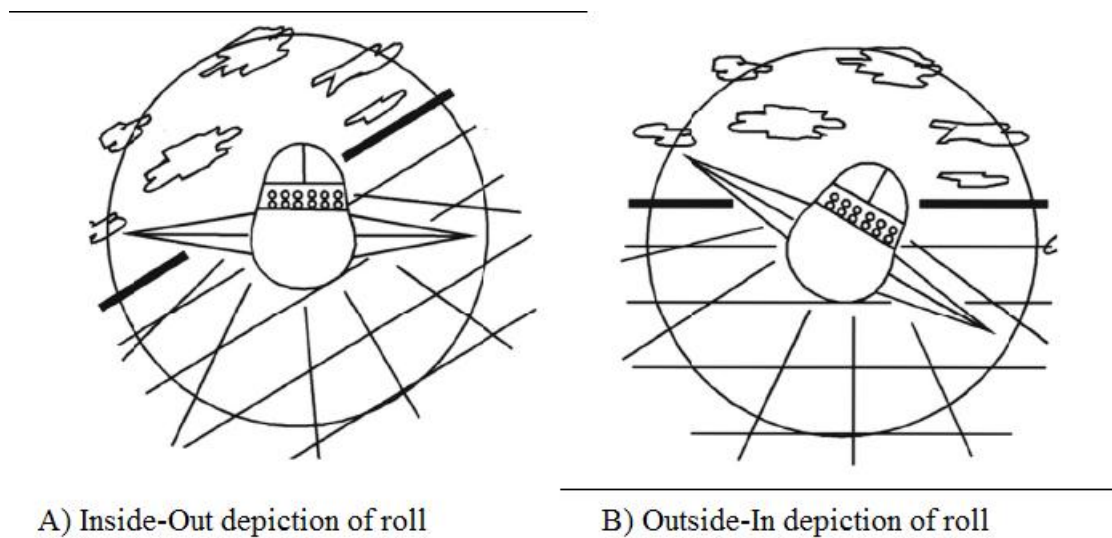
Figure 2.9 Aircraft attitude instrument

The Inside-out, FAA approved instrument (Figure 2.9.A), holds the image of the plane/pilot still and moves that earth's horizon moving in the distance. On the other hand, the Russian instrument (Figure 2.9.B) holds the earth still and moves the plane. Donovan and Triggs (2006) found that FAA approved instrument had significantly more "reverse control" errors, *i.e.*, turning the wrong direction and then having to correct that turn. It was suggested that the pilots tried to "move the horizon" to the wing orientation instead of moving the wing to the horizon.

One can recognize that each represents a valid method of presenting information to the pilots. However, if the pilots do not have the right frame of reference and/or point of view, the information must be transformed/interpreted by the pilot. The pilots can be trained using simulators or receive on-the-job training. If the pilot training is for a manned aircraft through simulators, the inside-out display matches the real cockpit perspective. However, if the pilot is flying by remote control using a stick control and watching the vehicle's movement, then the outside-in display more readily replicates how the student "thinks" about the maneuvers. When the display matches the way the operator thinks, it is a cognitive impedance match and the information is transferred from interface to user without distortion or dropout.

The historic weight of the inside-out instrumentation and training has made changing to the safer outside-in instrument nigh impossible. American pilots in exchange programs with Russian pilots have reported the difficulty of relearning the instrumentation.

The concept of cognitive impedance assumes the user interface (UI) is a communication channel that can be described by engineering descriptors. If this is the case, then a designer of user interfaces must understand the user/operator bandwidth and test for impedance mismatch as part of the interface design process. Obviously, signal encoding can affect the design. For instance, any video display must be formatted to match the engineered system and match the operator's internal visualization of the system. It might be easier for a dynamic display to adapt to the user's expertise, *i.e.*, shape the UI or even shape the user training. With a bidirectional communication, the UI can

41

adapt and be written in as a theoretical abstraction. For example, a context or concept map could drive the displayed UI.

2.1.4.4 Insertion of Information through Cognitively Sensitive User Interface/Methods

To improve the cognitive impedance match of additional network centric information, the information must be relevant to the current task at hand, otherwise it is a useless distraction or noise, and it must be presented in an easy to assimilate format that makes its value readily apparent. The visual thinking paradigm is a way to affect impedance matching. However, just presenting data graphically does not necessarily create impedance matching.

Within the Predator control system, three personnel are used to control the mission of the UAV. In particular, one of the job titles is the mission coordinator/controller (MC) with the responsibility of gathering, prioritizing and presenting external information to the other operators. This seems a strong argument by the Air Force for the requirement for an intelligent information channel that combines external information from diverse systems. The purpose of the MAGE project testbed is to aid the MC by presenting network centric information that is relevant, timely and presented in an easy to assimilate format for the UAV/UCAV pilot or sensor operator.

2.1.5 Situation Awareness

When one discusses human factors engineering and user interfaces in complex, dynamic systems, we are entering the domain of situation awareness. We say an operator

has situation awareness of the system when he/she has knowledge of all critical aspects of a controlled system. Endsley and Kiris (1995) formally defined situation awareness as the "perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and projection of their status in the near future". This definition of situation awareness suggested by Endsley and Kiris is echoed by many other researchers (Table 2.1).

Table 2.2 The definitions of Situation Awareness (SA) in complex and dynamic environments (Vidulich et al., 1994)

Definitions	Sources
Conscious awareness of actions within two mutually embedded four-dimensional envelopes.	Beriger & Hancock, 1998
The pilot's continuous awareness of self and aircraft in relation to the dynamic environment of flight, threats, and mission and the ability to forecast then execute tasks based on that perception.	Carroll, 1992
The ability to extract, integrate, access, and act upon task relevant information is a skilled behavior known as "situation awareness".	Companion, Corso & Kass, 1990
The accurate perception of the factors and conditions that affect an aircraft in its flight crew.	Edens, 1991
The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.	Endsley, 1990
The knowledge that results when attention is allocated to a zone of interest at a level of abstraction.	Fracker, 1988
The pilot's overall appreciation of his current 'world'.	Gibson & Garrett, 1992
One's ability to remain aware of everything that is happening at the same time and to integrate that sense of awareness into what one is doing at the moment.	Haines & Flateau, 1992
Where refers to spatial awareness… What characterizes identity awareness, or the pilot's knowledge of the presence of threats and their objectives, [as well as] engine status and flight performance parameters? Who is associated with responsibility, or automation awareness; that is, knowledge about 'who's in charge'. Finally, when signifies temporal awareness and addresses knowledge of events as the mission evolves.	Harwood, Barnett, & Wickens, 1988
The ability to envision the current and near-term disposition of friendly and enemy forces.	Masters, McTaggart, & Green, 1986
Awareness of conditions and threats in the immediate surroundings.	Morishige & Ratelle, 1985
The ability to maintain an accurate perception of the surrounding environment, both internal and external to the aircraft, as well as, identify problems and/or potential problems, recognize a need for action, note deviations in the mission, and maintain awareness of tasks performed.	Prince & Salas, 1993

Definitions	Sources
[Situational awareness] means that the pilot has to integrate understanding of factors that will contribute to the safe flying of the aircraft under normal or non-normal conditions.	Regal, Rogers, & Bouchek, 1988
Situation awareness refers to the ability to rapidly bring to consciousness those characteristics that evolve during flight.	Wickens, 1992
The pilot's knowledge about his surroundings in light of his mission goals.	Whitaker & Klein, 1988

Endsley's defined three levels of situation awareness: Level 1 SA is awareness or **perception**; Level 2 SA is **comprehension** within the context of the current task or operator goal; and finally, Level 3 SA reflects the ability of the operator to **predict** the future outcomes base on the current situation. A simplified model of SA in dynamic decision making inspired by Endsley (2006) is illustrated in Figure 2.10, but there are some key elements that have been moved for emphasis. Note that Level 3 SA depends on Level 2 SA, which in turn depends on Level 1 SA.

It should be noted, according to Klein (2009), that Situation Awareness is a measurement tool that takes no past information into account. In particular, it takes no user's domain experience into account. Further criticisms fall into two categories: 1) Are explicit measures necessary when more naturalistic techniques are available and appropriate (Durso, Bleckley et al. 2007) – is it long-term memory?; and 2) Then there are those that question the validity of the SA construct (Dekker and Woods 2002) – is this an artificial construct that is already addressed by research in attention?
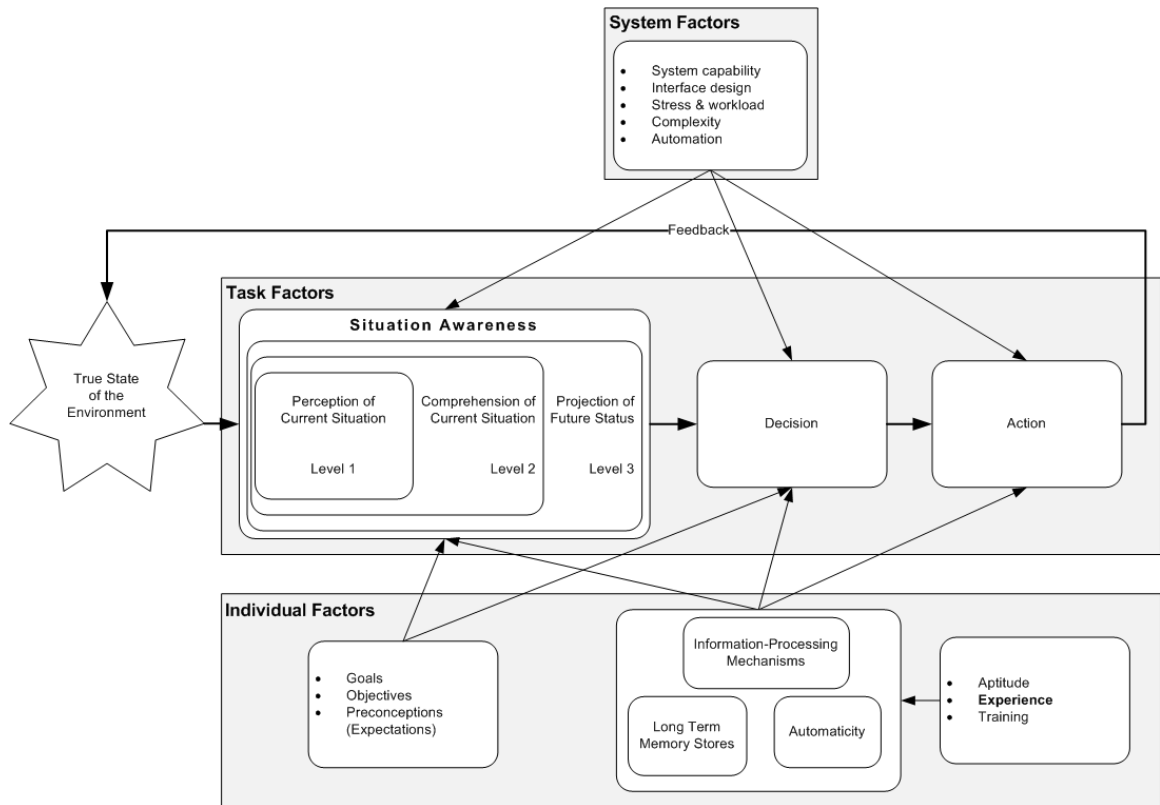
Figure 2.10 Model of Situation Awareness in a Dynamic Decision Making (derived from Endsley 2006)

Examining Figure 2.10 from top to bottom, the **System Factors** are those factors inherent to the task domain or designed into the task/user interface. For example, operators monitoring multiple aircraft within a small domestic market such as Dayton, OH (1.3 million total passengers during 2006 and nestled on a very sparse 4,500! acres http://www.daytonairport.com/index.htm) is quite different from operators monitoring a large international market such as the Ronald Reagan International Airport, Washington, DC with 18.7 million total passengers during 2006 (10.2 times more passengers than Dayton!) and a very dense 840 acres (only 733 acres above water!) http://www.metwashairports.com/reagan/about_reagan_national/air_traffic_statistics_2]). The mental stress of the air traffic controllers in Dayton International Airport can be

assumed to be much less than Ronald Reagan International Airport. The scale and complexity of the software and hardware for each market must be much different. The automation and decision support included in the systems will be of different complexities.

The **Task Factors** involve an operator feedback loop (very similar to the Boyd's (1996) OODA loop) that interprets elements of the real environment to construct a mental map, called the situation awareness, as the basis of decisions and actions that follow. If the situation awareness is not complete (enough), the decision and the action can be faulty. The feedback loop allows the operator to self-correct if the action does not fit the predicted outcome.

Finally, the **Individual Factors** can bias or shade the situation awareness of the operator. Because the operator's goals and objectives may differ slightly from the optimal solution's objective, a first time operator may find an acceptable goal to be not crashing, whereas the mission statement may require a full surveillance of the target zone. Furthermore, an individual operator's personality, background, and resourcefulness may shade the preconceptions of what is or is not acceptable within the mission goal parameters. Each operator will have their own level of aptitude, experience and training. It is hoped that training will weed out the inept, but this is not guaranteed. Furthermore, additional training is often substituted for actual experience, but it is precisely the combination of aptitude, experience and training that provide the support for the skills (information processing skills), long term memory stores, and automaticity. (Automaticity is "performing without thinking", for example, the automatic muscle

memory that allows a driver to automatically drive home without thinking about the route. Whole segments of a trip can be "blanked out" by the driver, with no understanding of the details of how they got from point A to point B.)

Of these three factors, situation awareness of the system is experienced in the Task Factors. The other two factors influence or bias the interpretation of existing information of the current situation, but the actual actions based on the operator's understanding of the situation is in the Task Factors. Next, the individual levels of SA are examined.

2.1.5.1   Level 1 Situation Awareness (Level 1 SA)

The "basic building block" of all levels of SA is Level 1 SA. Level 1 SA is the recognition that something in the environment needs attention. This is the most basic of all situation awareness levels, but failure to recognize a problem in the environment can have horrendous repercussions. The processing of Level 1 SA is "bottom-up"[7], i.e., the scanning of all the data to find any unusual data points.

---

[7] "There are two basic modes of processing. 'Bottom-up processing,' also termed 'data-driven processing,' is processing initiated when data are bound to variables in bottom level subschemata that move upward to activate the higher level schemata in which the subschemata are embedded. 'Top-down processing,' also called 'conceptually driven processing,' is processing initiated when top level schemata activate embedded subschemata in the expectation that these subschemata will fit the data … Data-driven processing moves from part to whole, and conceptually driven processing moves from whole to part … Data-driven processing is subconscious, automatic, and guided by the principle that 'all the data must be accounted for,' while conceptually driven processing is conscious, purposive, and guided by high level plans and goals."
    Casson, R W (1983). "Schemata in Cognitive Anthropology." Annual Review of Anthropology **12**(1): 429-462.

Approximately 76.3% of accidents reviewed by Jones and Endsley (1996) for the National Transportation Safety Board (NTSB) were directly attributed to failure to perceive needed information. It thus follows that the level of SA to which user interfaces need to focus is Level 1 SA and the critical cues of the domain need to be emphasized by the designer to aid the operator.

How does one design a Level 1 SA UI? Endsley and Kiris (1995) posit that critical cues are perceived in the environment when individual elements are classified using pattern-matching prototypes; this in turn activates corresponding mental models in long term memory. What Endsley's theory of SA does not address is an approach to create visual displays that present critical information cues tailored to the operators goals and objectives. SA is enhanced when the critical cues attract the operator's attention, are tailored to the task (within constraints), and are related to the operator goals. By utilizing the operator's innate 2D pattern recognition abilities (sorting relative area), the visual thinking display addressed Level 1 SA requirements, and thus enhance overall task performance.

2.1.5.2   Level 2 Situation Awareness (Level 2 SA)

Nested within the Level 1 SA (perception) is Level 2 SA. This level recognizes the need for the operator to contextualize the perceived situation. To contextualize, the operator accesses "schemata or knowledge stored in long-term memory (Rumelhart, 1984) which are activated by recognized patterns in incoming data" (Endsley 1988). Level 2 SA is "top-down" processing, i.e., goal directed. An operator has a set of goals that directs their attention to specific, relevant information in the environment.

Level 1 SA ("bottom-up") and Level 2 SA ("top-down") work in combination and in parallel with each other in a means-ends process that, if working properly, quickly zooms in on pertinent information. Each level depends on the other to fully develop the operator's situation awareness. When working together, the perceived data and the operator goals afford meaning and significance, transforming the disconnected data into relevant information.[8]

How important is Level 2 SA? Approximately 20.3% of accidents reviewed by Jones and Endsley (1996) for the National Transportation Safety Board (NTSB) were directly attributed to failure to Level 2 SA. This means the operator was able to detect or perceive that a problem existed, but was not able to recognize its importance or meaning within the context of the operator's mission. That is, the perceived problem's importance was down played.

According to Endsley and Kiris (1995), this combination of perception and contextualizing "activate the appropriate goals and models", i.e., it triggers a cognitive schemata or pattern. Connecting the appropriate situation pattern to the current situation allows the operator to access long-term working memory for suitable mental models. Activating the mental models is a "bottle-neck" within the information processing system of complex, dynamic systems (Endsley 1988). Designing a user interface that facilitates this connection process would assist situation awareness: the design has to be domain

---

[8] The use of the terms data versus information is intentional. From a strictly cognitive science viewpoint: data + meaning → information. For example, 12, 3, 2 are data, but if you include the context (meaning) of being numbers within Christmas song titles, they become information ("Twelve Days of Christmas", "We Three Kings of Orient Are", "All I want for Christmas is my Two Front Teeth").

specific to capture the significant data in the environment; and help the operator make the connections by utilizing the operator's cognitive strengths.
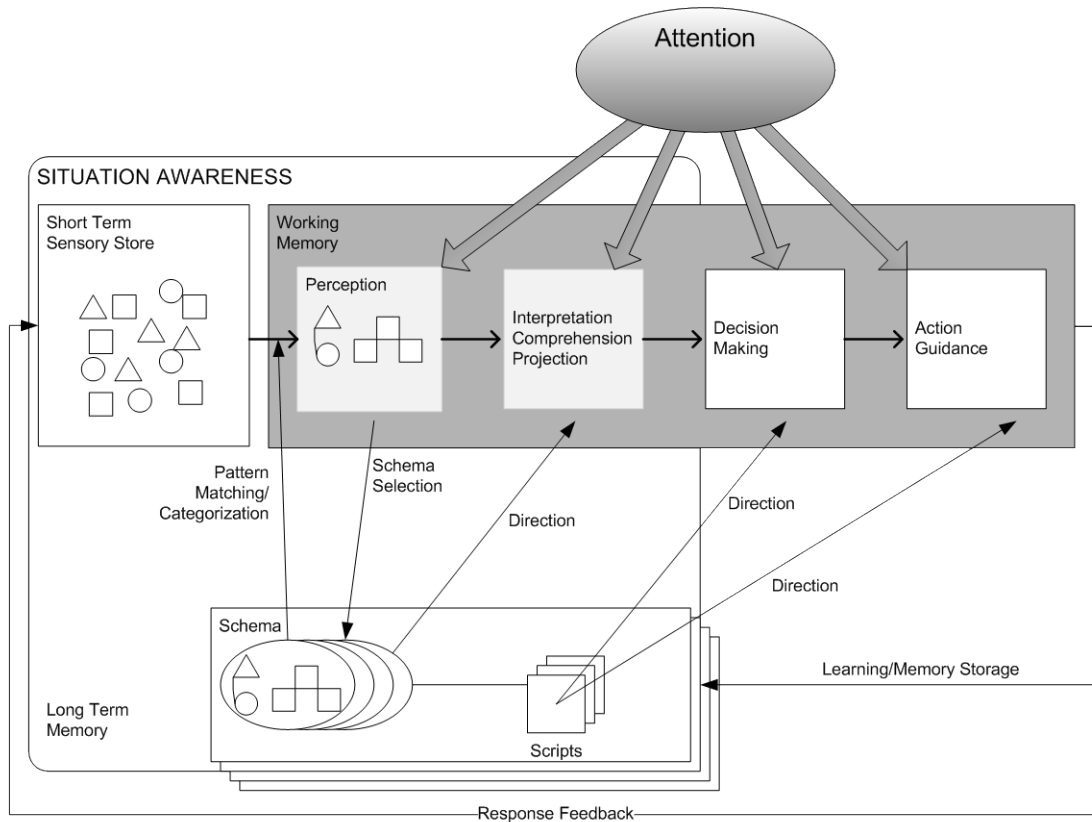
Figure 2.11 Cognitive processes influencing the formation of SA (Endsley and Kiris 1995)

This active combining of Level 1 SA and Level 2 SA is situation assessment (Endsley and Kiris 1995), where "internal cognitive constructs such as attention, pattern matching, and long-term working memory influence what data in the environment" is attended to, and what the final "snapshot" of SA encompasses (Figure 2.11). Additionally, experience as a major external influence from the environment. (Of course, SA is always fluid/dynamic and never static nor complete. The "final SA" is the incomplete SA used to make decisions.)

To summarize the above discussion, situation awareness is not passive, and most definitely situation assessment is a very cognitively active. The activities driving situation assessment are separate from the resulting constructed situation awareness. As Tenney et al. (1992) states,

> [t]he state of awareness with respect to information and knowledge is the product. The process, in contrast, involves an active and dynamic series of cognitive activities. Maintenance of situation awareness is not easy because the process requires mental resources that may be in competition with ongoing task performance. The information gathering activities that contribute to situation awareness therefore may heighten workload momentarily. However, a principal benefit of achieving situation awareness is that the operator or crewmember is prepared to deal with upcoming events such that the extreme surges in workload that can occur in unexpected circumstances are avoided. [p.2-3]

2.1.5.3   Level 3 Situation Awareness (Level 3 SA)

Finally, the final level of situation awareness is Level 3 SA, projection. Level 3 SA is the result of the active cognitive mechanisms (attention, pattern matching, and long-term working memory) behind Level 1 SA (perception) and Level 2 SA (context). The perception and fusion of external data from the environment (Level 1 SA) along with information culled and guided by the operator's goals and mental models allow the operator to predict or project potential future states of the system (Level 3 SA).

Recall from the previous two sections, accidents reviewed by Jones and Endsley (1996) for the National Transportation Safety Board (NTSB) were directly allocated to 76.3% (Level 1 SA) and 20.3% (Level 2 SA). In other words, accidents directly attributable to the Level 1 SA and Level 2 SA were 96.6% of the total accidents and only 3.3% was attributable to Level 3 SA. One way of explaining this is that once the problem is identified and placed in context, most pilots are able to respond satisfactory to the

existing problem. Therefore, designing a user interface just to address Level 3 SA does not have a large return of investment. For that reason, this research will be designing the user interface towards Level 1 SA and Level 2 SA. Improvements in the Level 3A will be considered icing on the cake.

2.1.5.4 Situation Awareness Assessment Tools – SAGAT and SPAM

After this long discussion with regards to situation awareness, one would naturally ask: "What does this give me?" and "How do I measure SA?" The SA theory allows one to directly measure complex dynamic situations. Thus, comparing two different user interfaces can be achieved directly by comparing the operator's SA for each interface.

With regards to the second question, one must next decide whether the subjects are required to memorize the situation or are they allowed to "look up" the answer. If the first, then the subject should be well versed in the domain! The operator must be able to mentally reconstruct the supplied information to reach a conclusion. Since our subjects for the first two experiments were not subject matter experts, this did not seem feasible. If the latter, then the SA measures the subjects ability to locate the correct information! This seems to more accurately reflect real world complex domains. Thus, the SA assessment tools considered for this experiment were SAGAT and SPAM, two very interesting assessment methods. (There is a veritable alphabet soup of assessment methods, but these two are of particular interest within the operator assessment domain according to the European Organization for the Safety of Air Navigation (EOSAN).)

"The Situation Awareness Global Assessment Technique (SAGAT) is a global tool developed to assess SA across all of its elements based on a comprehensive assessment of operator SA requirements" (Endsley and Kiris 1995).   Endsley's assessment technique requires the stopping of the experiment while hiding the information on the display.   The Situation Present Assessment Measure, or SPAM (Durso, Rawson, & Girotto, 2007), assesses the speed of accessing information from a nonblanked display and provides a more sensitive, continuously distributed (time) measure that will be less likely (than SAGAT) to be at floor levels because of memory decay.   Although Endsley prefers the SAGAT assessment tool, this research followed the recommendations of the EOSAN for measuring SA in the UAV domain and used SPAM, precisely because the subjects were inexperienced.

2.1.6 Separate, Integral and Configural Displays

Finally, one can describe this display with respect to display design organization. Any user interface performs two functions: it presents the important state information (as envisioned by the designers) to the user and it enables the user to perform tasks to accomplish their goals. An interface should extract the critical features of the problem space to enable users to achieve their goals. Furthermore, "the interfaces control-display relationship should be consistent with human perceptual and cognitive abilities so that effective control of the design process can be achieved" (Rothrock, Barron et al. 2006). Another view of this presented by Edlund and Lewis (1995) states that any design methodology must:

1. Make a relevant process easy to discriminate
2. The process behavior (changes in state) intelligible
3. If mapping between display states, behaviors in the process being controlled "easy" to follow
4. Do all the above simultaneously, within a single representation

This is very difficult and has had much research time expended on designing the visual representations (Pomerantz 1986; Barnett and Wickens 1988; Wickens and Carswell 1995; Jenkins 2007).

2.1.6.1   The Gulf of Evaluation

Pomerantz (1986) proposed three relationships among visual stimuli: separable, integral and configural. A separable relationship has no interaction between dimensions, such as sound volume does not affect the perception of length. At the opposite extent, an integral relationship cannot separate the relationship between dimensions, such as mass and gravity combine to form weight. An intermediate configural relationship can be defined by the independence of the dimensions that has an emergence of a new property, such as the relative sizes of independent population bubble graphs. There may not be an explicit ordering of the data on the graph, but the observer can identify an implicit ordering based on the size of the bubbles – this sort ability is the new emergent feature.

According to Wickens and Carswell (1995) the *Proximity Compatibility Principle* (PCP) "specifies that displays relevant to a common task or mental operation should be rendered close together in perceptual space." This defines the relationships between task demands and the graphical form of a display (separable, integral, configural).

The PCP's "close together in perceptual space" means both display (physical) proximity and mental proximity. To have display proximity nearness one needs: spatial proximity (physical distance), chromatic proximity (same or different color), code homogeneity (coding properties to be similar or different), and geometric form (integral or configural versus separable displays).

On the other hand, mental proximity, meaning the "extent to which information from the various sources in a display must be considered together to accomplish a task" (Rothrock, Barron et al. 2006), has three categories: integrative processing (highest proximity), nonintegrative processing (intermediate proximity), and independent processing (lowest proximity). Integrative processing from multiple sources must be explicitly combined (e.g., probability of survival requires information about SAM sites, enemy aircraft and enemy positions). Nonintegrative processing similar features can be categorized (e.g., oil- and engine-temperatures). On the other hand, independent processing has no interaction between sources (e.g., distance-to-base and air-temperature). (However, to calculate time-to-return-to-base may require distance-to-base and air-temperature.)

The Gulf of Evaluation (GoE) is the "nearness" of the display proximity and the mental proximity. An efficient interaction has a small GoE and occurs when the display proximity matches the mental proximity, creating a cognitive congruence. When these do not match, there exists a large GoE or cognitive dissonance. Performance on integrated tasks (high mental proximity) is predicted to be facilitated by displays that have high perceptual proximity (integral or configural displays), i.e., analog gauges are

more conducive to driving a car. By the same token, performance on focused tasks (low mental proximity) is predicted to be facilitated by displays that have low perceptual proximity (e.g., text tables are good for baseball batting statistics).

2.1.6.2 The Gulf of Execution

PCP may constrain displays so that efficient interaction can occur, it does not, however, specify the form. Display control literature has developed general design guidelines called the *principles of control-display compatibility* (Wickens, 1992), that stipulate:

- The spatial arrangement of controls should allow users to easily tell which control is used.
- The indicator of a display should move in the same direction as its control.
- The layout of the operational method of controls should be consistent with expectations of the user population.
- The direction in which a part moves on the display should be consistent with user expectations.

These principles serve as a guide that uses user expectations to guide the display design choices. For example, "bigger is better" might translate to a bar graph size indicating the relative merits of that dimension. When these expectations are not met, there exists a mental dissonance and the Gulf of Execution (GoE) is said to be larger.

These principles have been incorporated into the design of the visual thinking display artifacts and the entire suite of displays for this research. The text tables were modified to incorporate color as a visual cue to out of tolerance, and thus become configural displays of the most basic type. The bar graph display was chosen as a naïve design attempt as a graphical user interface. The bar graphs were grouped by dimension (FoM) to create a display proximity advantage for the dimension (number of SAM sites).

However, by definition of the bar chart form, this in turn presented a poor display proximity for the by UAV tasks (Which UAV had a better chance of survival?). This mental proximity disjointedness is inherent to the bar graph feature set – if the bar graph is displayed along the UAV axis instead of the FoM axis, the mental proximity problem would just be reversed. The mission decision aide (the configural visual thinking display for experiment 1) incorporated the same out of tolerance color cue, had the within UAV display proximity (the widget itself) and across dimension display proximity (adjoining widgets had dimensions represented in the same relative positions, making comparisons relatively easy).

2.2 Domain

The use of unmanned vehicles in the military and Homeland Security applications are specifically for those missions that are too long, dangerous, or dirty to waste human resources. When a mission has a long duration, the human operator's attention may wane; resulting is in a less than stellar mission result. In dangerous missions, such as surveillance over enemy territory or rescue missions in toxic environments, the use of unmanned vehicles lowers human injury and death rates. Finally, missions that are uncomfortable can be a distraction to the human operator, possibly causing the failure of mission objectives.

As a final argument for the use of UAVs, let us examine the economics of using and crewing UAV systems. UAVs are simpler and easier to operate than manned aircraft; hence the cost of training is much less. For instance, training a rated pilot costs over one million dollars and incurs a yearly salary of $80-$100K. Compare that to

training costs of UAV operators from the enlisted ranks ($30K), with salaries closer to $30K per year. Furthermore, since there is no onboard crew, no crew safety training nor life support systems need to be integrated into the UAV systems. Finally, the control systems can be made simpler, bringing down the cost further and simplifying design.

The U.S. military is interested in applying single operator/supervisor control of multiple UAVs to their unmanned combat aerial vehicle (UCAV) missions; in particular, (1) search, (2) search and rescue (SAR), (3) search and destroy (SAD), and (4) suppression of enemy air defenses (SEAD) missions (SRA International Inc. 2005). SEAD "is the activity that neutralizes, destroys, or suppresses enemy [Air Defense (AD)] systems in a specific area by physical attack and [Electronic Warfare (EW)] to enable [Tactical Air] operations to be successfully conducted. It increases the probability of success and reduces the loss of friendly air power" (Army 2000). The SEAD mission has two forms: non-lethal (disruptive suppression) and lethal (destructive suppression). Disruptive suppression complements destructive suppression, and is best used to: (1) degrade jammable threats; (2) assist destructive ground-based and airborne suppression systems in suppressing surface-to-air defense systems; (3) temporarily degrade or neutralize enemy AD systems when destruction is not possible or feasible; and (4) sustain suppression effects achieved by destruction, once the threats have been reduced to levels commensurate with the objective.

If we compare the two varieties of SEAD mission in a communications scenario, the differences become clear. The non-lethal SEAD mission employs UAVs that use electronic countermeasures to jam communication channels, thereby suppressing the

enemy command and control. Non-lethal SEAD can also use force to temporarily damage communications towers, without obliterating the whole communications infrastructure. On the other hand, lethal SEAD destroys the enemy communication infrastructure using any military force deemed necessary, including armed UAVs. The intent is to inflict so much destruction that alternate means will be needed to communicate within the military and civilian societies for the foreseeable future. Recovering from a non-lethal SEAD mission is relatively painless compared to the rebuilding required if the lethal SEAD mission achieves its goals. As the Army Field-Manual 1-114 (2000) points out, lethal and non-lethal SEAD methods can be used in conjunction with each other. Each of these SEAD missions is currently used by the military (individually and collectively). This research addresses only the non-lethal SEAD mission.

2.2.1 Control of UAVs

Controlling UAVs falls under three categories: (1) individual, (2) small team, and (3) swarm size. Each of these categories requires different assumptions and levels of control. Currently, the standard control paradigm is flying individual UAVs by a small group of individuals. The UAVs do not need to be autonomous and the control mechanism can model operator control. Small groups require a more sophisticated, semi-autonomous UAV and transfer the control mechanism towards a supervisory role. However, the "pilot" may take operator level control of individual UAVs. On the other hand, swarm control requires a smart, autonomous UAV with a nearly pure supervisory control model.

The Predator family of UAVs is the leading deployed production UAV system. During the 1990's, the usefulness of the Predator as a surveillance platform was proven very effective. The Predator UAV started as an operator controlled vehicle, but is slowly evolving towards a semi-autonomous combat workhorse. Since this research required an existing UAV system to test its theories, the Predator's evolution of purpose and sophistication fit those requirements.

Although swarm control is an interesting topic by itself, this research only examined the controlling of individual UAVs and small teams of UAVs.

2.2.1.1 Control of Individual UAVs

Individual UAVs can be controlled by pilot operators in a manner similar to how a pilot controls a manned combat aircraft, i.e., stick and rudder. However, since the operator is situationally removed from the UAV and the feedback is drastically limited, forming a technological anomie, as experience has shown that the piloted aircraft control model is not the correct model with respect to UAVs. Experience has changed the control model to a "point and click" mouse driven operator interface.

The primary consumers of UAVs, the military and the Department of Homeland Security (DHS), discussed the control of UAVs during the CERI 2nd Annual Human Factors of UAVs (2005) sponsored workshop. Question and answer sessions were held with current UAV operators from the United States Air Force and Army along with a representative of the DHS and the Federal Aviation Administration (FAA). The following information was gleaned from those discussions.

One of the major concerns for the operator is situational awareness because of the lack of sensory feedback. Each of the military branches uses their own resources and philosophies to try to address this problem. Each branch had a representative discuss these military philosophies at the Cognitive Engineering Research Institute (CERI) 2nd Workshop.

Currently, in the United States Air Force, each Predator UAV requires three persons to complete a mission: an operator, a sensor operator (SO), and a mission controller/coordinator (MC). The Predator is a large UAV, approximately one third of the size of an F-16 fighter aircraft. The corporate mindset of the Air Force considers the UAV operator to be equivalent to a rated officer; this means the operator is an officer who has gone through the same initial training as any pilot or a rated navigator who has passed instrument flight ground school training. Since the Air Force considers the ability to "place one's self into a remote vehicle" difficult, it has stated that it expects a pilot may crash at least one UAV during training and still remain a viable candidate for pilot operator (Gunter and Lytle 2005). Furthermore, the operators may be literally located half way around the world from the battlefield; for example, while the UAVs flying in Iraq II were launched and landed by pilots stationed in Iraq, the missions were flown by operators based in Nevada. Although difficult, projecting oneself into the remote vehicle is seen as essential to effective operator situation awareness.

On the other hand, The United States Navy requires only two rated officers as operators for their UAVs. The navigator and weapons specialist positions are combined (Cummings and Guerlain 2004). The Navy's UAV interest seems directed more towards

intelligent munitions that can be launched from sea and redirected afterwards. Retrieval of the weapons system does not seem to be an important issue. The naval intelligent munitions can be launched from an aircraft carrier, while the operator can be located on the same ship, a different ship, or on a naval base on land. Once again, the operator's ability to project themselves into the remote vehicle's situation is considered essential to the success of the mission.

The third military branch, the United States Army, uses enlisted and warrant officer personnel to operate their rotary and fixed wing UAVs. When choosing personnel for training, the army selects the enlisted personnel with experience using computer game and/or sophisticated model airplanes. The Army expects the knowledge gained in immersive gaming and model airplane flying to translate into the operator being able to project themselves into the UAV's remote situation.

2.2.1.2   Multiple UAV Control

As Cummings (2004) pointed out, "while currently [UAVs] require relatively concentrated input for flight control, in the future, it is likely that the human role for direct flight control will diminish and the need for supervisory control, to include higher-level cognitive reasoning, will become much more substantial." In other words, in order for the services to progress from the individual UAVs to multiple UAVs, the individual UAV is required to become more autonomous. This changes the control model from an operator/pilot control paradigm to a supervisor control paradigm. The supervisor control paradigm presents a lower human risk and cost than traditional pilot-controlled aircraft and it is a lower cost alternative to single UAV operator-controlled aircraft.

Recent research examining autonomous pack control issues have examined heterogeneous teams (Tambe 1998), homogeneous wolf packs (Lewis, Polvichai et al. 2005) and autonomous wingman (Cummings and Morales 2005). Tambe's (1998) research examined heterogeneous UAV teams, in which each UAV had a unique role and explicit responsibilities – commander, scout and attacker. Each role had explicit rules; had to have all failures explicitly accounted for; and explicitly defined rules for recovery. This strict enforcement of roles is a very brittle, complicated system that does not easily scale up to larger groupings. Furthermore, the teams used an "explicit model of teamwork", known as the *joint intentions* framework. The joint intentions framework creates an intention or commitment by the team to perform a task. To create the commitment, all teammates must mutually believe: (1) that the task has not been performed, (2) the task needs to be performed, and (3) that until the status of the task changes to be mutually known to all teammates to be *achieved*, *unachievable* or *irrelevant*, it remains a goal. If one team member finds that the task becomes achieved, unachievable or irrelevant, then this team member makes its new task to be that of changing the team's mutual belief to reflect its own beliefs.

When one compares this approach with Lewis, Polvichai et al. (2005), teams are referred to as packs; the packs are homogeneous – utilizing intelligent munitions; and modeled after wolf pack attacking methods. Most importantly, in Lewis' approach, there are no unique roles. For instance, in a search and destroy, each UAV is initially assigned a search role. If a UAV finds a target it calls for reinforcements to form a pack. Once it is formed, the pack may attack *en masse* or sequentially with a lone UAV always held in reserve to perform battle damage assessment. If the target is of high enough priority, the

arrival of a second UAV can initiate the attack on the target by the first UAV. If the first UAV does not destroy the target, then the second UAV waits for a third's arrival before proceeding to attack, and so on….

The autonomous wingman (Cummings and Morales 2005) typically uses the UAVs as defensive weapons and battle damage assessment recorders. The UAVs assigned the "wingman" role in a tactical air formation are controlled remotely by the lead manned aircraft. Who, what, where, when, and why the pilot's commands are executed is of intense interest to the researchers of the autonomous wingman paradigm. As Cummings and Morales (2005) point out, "[p]reliminary research suggests that without higher levels of autonomy and a shift from management-by-consent to management-by-exception control strategies, the workload of pilots controlling UAVs in-flight, especially single seat pilots, will be too high."

2.2.2 Autonomous Agents

How autonomous is autonomous enough? If we assume that technologies already available to larger aircraft (automatic take-off and landings, collision avoidance, etc.) will soon be available to UAVs, then the argument can be made that the current or near term UAVs are autonomous enough. These leading edge UAVs can be modeled as an autonomous intelligent agent and any improvements to the current technology can also be modeled.

Table 2.3 Agent criteria and rationale

Ferber (1999) defines an autonomous agent is a biological, mechanical or virtual entity	Rationale
that is capable of acting in an environment;	Environment affects decisions each UAV makes, pop-up entities can affect the UAV's route
that can communicate directly [or indirectly] with other agents;	Communicate directly and indirectly (through other agents) with other agents
that is driven by a set of tendencies (has autonomy);	Target detection affects tendencies
which possesses resources of its own;	Has weapons and fuel
that is capable of perceiving its environment (although limited);	Perception of physical and virtual stigmatic information
that has only a partial representation of this environment;	Local view that can be augmented with neighboring agents views
which possesses skills and can offer services;	Can search for or attack targets
that may be able to reproduce itself; and	
whose behavior tends towards satisfying its objectives, taking account of the resources and skills available to it and depending on its perception, its representations in the communications it receives.	Each agent has a specific role and mission

In preparation of the UAV hardware technology catching up with the requirements of autonomy, the Multiple-UAV Agency (MAGE) simulation testbed was created. Table 2.3 shows how the MAGE UAV agents fit the agent programming criteria as defined by Ferber (1999) and the rationale applied to show how the UAVs in the MAGE simulation fit these criteria. The MAGE environment is a distributed heterogeneous multi-agent simulation testbed using a flexible and portable interactive simulation infrastructure designed to be extensible. MAGE is a research testbed that permits exploring different facets of controlling multiple UAVs, such as, network centric information synthesis, supervisory control, path planning, multi-sensor data fusion, etc.

However, having an agent in an environment is not enough. Agents have only a partial, local representation of their environment. There is no global awareness of the evolving environment (situation); this exactly mimics how it is in human society. To overcome this weakness, humans form groups and share local information to create a

more global awareness. This awareness does not guarantee omniscience, but rather less situation ignorance. The equivalent in the agent paradigm is the multi-agent system (MAS). Table 2.4 illustrates how the simulation within the MAGE structure will fulfill the MAS criteria.

Table 2.4 Simulation swarm criteria and rationale

Ferber (1999) defines a MAS application as requiring	Rationale of MAGE as MAS
an environment, E	Simulated geographic location, including bases, no-fly zones, etc.
a set of situated objects, O	MAGE objects include targets, bases, UAVs, "red/blue forces", etc.
an assembly of agents, A, (a subset of objects)	Some of the simulation agents are the UAVs agents and the target agents
an assembly of relations R, that link the objects (and agents)	A UAV can communicate capabilities, mission, etc.
an assembly of operations, Op, allows the agents to act on the objects; and	Each UAV search, destroy, perceive, etc.
operators whose task is to represent the application and react to these operations.	Decision Support Interface (DSI) agent, Wizard of OZ (WOOZ) agent, etc.

Based on these two lists of criteria, this simulation fits comfortably into the agent-based programming and MAS paradigms. However, the "intelligence" of the individual agents may be programmatically internal or external to the agent, *i.e.*, the MAGE software contains legacy software that does not necessarily contain fully encapsulated agents.

2.2.3 Humans in Complex Systems

Sheridan (1997) asserted that there are many alternatives to the human/machine interface, the extremes of the spectrum being technology-centered and human-centered paradigms. The technology-centered paradigm insists "everything that can be automated should be automated," while the human-centered approach believes one should "allocate

to the human the tasks best suited to the human, allocate to the automation the tasks best suited to it." Within human-centered automation exists a wide spectrum of alternatives: static or dynamic allocation; by function or by mission.

If we contrast the goals of technological automation to human-centered automation; we find that the technological automation is trying to take the human out of the system (presumably, because it is the weak link). As Hildebrandt and Harrison (2002) precisely state, "automation promises to extend or support human performance, to compensate for human performance deficits, to relieve the human of routine tasks, *or to replace the human altogether*." On the other hand, the human-centered automation goal creates human/machine systems in which the operator retains control in well-defined tasks such that the performance of both the machine and human is optimized (Hildebrandt and Harrison 2002). In other words, the human should be intimately involved in the process without being overwhelmed. It also means the human and machine need to work as a cohesive unit, and not as disjoint components.

In dynamic complex systems, the human interaction in the system has two control methods at opposite extremes of the autonomy spectrum: human-in-the-loop (HIL) and human-out-of-the-loop (HOL). The HIL operator has more manual control over its system and, hence, is more active in the decision process; while the HOL operator has less control (since the system tends to run itself) and so this operator tends to be more passive. The HIL operator tends to remain more vigilant and less complacent, since they are actively assisting in the decision making process. The active involvement of the operator improves the operator's situational awareness (SA) (Ruff 2000). HIL operators

typically have faster response time and more accurate failure detection performance than HOL (Hildebrandt and Harrison 2003).

When the operator is performing in HOL mode, then the individual UAV agent being controlled must be fully autonomous. Endsley and Kiris (1995) argue the loss of manual skills is a major concern for HOL operation. After interviewing eleven fighter pilots, Rouse, Geddes et al. (1987-1988) found "... there is a clear consensus [among the pilots] that the pilot should be in charge." This clearly makes a reasonable argument for lower levels of automation so that one can achieve HIL operator efficiency and maintain SA. These human/agent systems are replacements for currently existing systems; that is, a pilot is replaced by an operator and a manned aircraft is replaced by an unmanned aerial vehicle. Notice the one-to-one correspondence with pilot/operator and plane/UAV. The previous research assumes that the system design must mimic current job divisions and methodologies. This argument is not necessarily sound for the emerging multiple UAV systems research.

However, with regards to multiple UAV control, the loss of manual skills should not be the reason for overloading the operator with information. Multiple UAVs must be able to be controlled separately and as a unit. Two facts should seem obvious: 1) one person cannot (simultaneously) fly multiple UAVs manually, and 2) a completely autonomous UAV will not be trusted. One must find the right level of automation to assure the best SA for the operator/supervisor. As Ruff (2000) points out, even automation intended to function autonomously will occasionally require operator intervention.

It thus follows that we must find a level of automation, which is dynamic, somewhere between fully autonomous and manual.

2.2.4 Levels of Automation

Different levels of automation can be introduced into a decision support system from fully automated, which leaves the supervisor/operator out of the decision process, to the computer making no decisions and thus forcing the operator to make all of the decisions. Modern supervisory control inserts a computer to track progress of a complex dynamic task (see Figure 2.12)
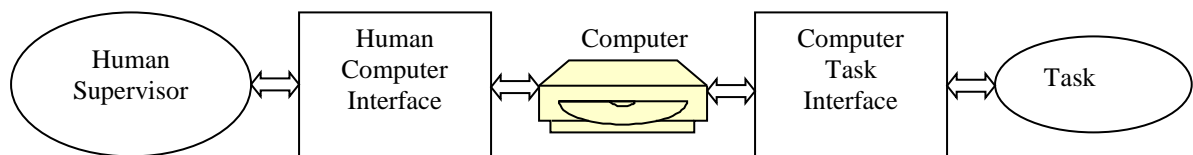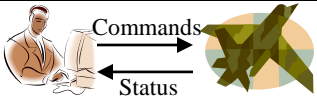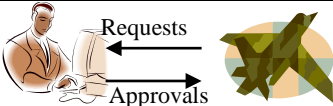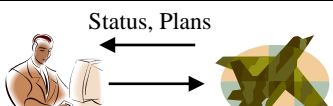
Figure 2.12 Human supervisory control – adapted from (Sheridan 1997)

Parasuraman, Sheridan et al. (2000) suggests ten levels of automation (LOA). These levels of automation can be categorized into three potential LOA paradigms: (1) manual control [level 10], (2) management-by-consent [levels 5-9], and (3) management-by-exception [levels 1-4]. In manual control, automation is inactive until explicitly activated by the operator, when the operator gives up control. In management-by-consent, the automatic problem solver proposes actions and requires explicit operator approval to proceed. Management-by-exception allows the automatic problem solver to act without requesting approval from the operator and will not act only when explicitly commanded by the operator.

Table 2.5 Different operator control paradigms provide different sets of strengths and challenges to C2 interface designers. As always, designers should examine potential methods for combining the aspects desired into hybrid approaches. (Lewis, Polvichai et al. 2005)

Direct Control	Management by Consent	Management by Exception
Commands → / ← Status	← Requests / Approvals →	Status, Plans ← / Overrides →
• Operator does all decision making and information processing • Requires operator to constantly attend to vehicle • High workload	• Vehicle performs planning and sends plan to operator for approval • Vehicle performs no action without obtaining operator approval • Operator highly interruption-driven • Operator must react quickly to ensure vehicle safety for time critical actions • Moderate workload	• Vehicle performs planning, sends plan to operator, begins execution • Operator has ability to override vehicle actions, plans • Operator must maintain awareness of situation • Requires high degree of intelligence, autonomy for vehicle • Low workload

Franke, Zaychik et al. (2005) illustrated the different operator control paradigms explicitly (see Table 2.5). The military branches currently use direct control for individual UAVs. Pack control require management-by-consent and management-by-exception. The level or emphasis is related to the autonomy of the UAVs.

Ruff (2000) found that management-by-consent was the most appropriate operator control mechanism for a small number of UAVs, usually considered a pack level (four or less UAVs).

2.2.5 Human Role in Multiple UAV Control

Cummings (2004) suggests that within the supervisory levels of automation proposed by Sheridan (1997) an additional layer describing the inter-vehicle

communications levels of automation needs to be addressed to guarantee situational awareness for the supervisor (See Figure 2.12).

Table 2.6 Operator and supervisor characteristics

	Operator	**Supervisor**
Role	Tactical/Operational	Strategic
Level of Automation	Low-to-High	Medium-to-High
Minimum Level of Autonomy	Minimum network	Medium network
Maximum Size	Small (2-4)	Small-to-Medium (4-12)
Control paradigm	Direct control	Guidance only

As Table 2.6 illustrates, there are two possible roles for human control in complex dynamic systems: operator and supervisor. The operator's role is predisposed towards an operational approach with intervention behavior similar to an intermittent correction servo (Scerri, Xu et al. 2004; Lewis, Polvichai et al. 2005) , while the supervisor's role tends towards strategic planning, such as assigning a computer to solve a problem, and then only intervening when computer makes a mistake or requires assistance.

The role of the human evolving from an operator to a supervisor is illustrated in Figure 2.13. The nodes represent entities involved in the performance of a high level mission. The arrows represent implicit and explicit exchange of information over time throughout the mission. The Human Computer Interface is any controls and displays necessary to interact with the computer. The Computer Task Interface is any necessary actuators and sensors the computer needs to interact with its environment.

a) Operator Control

b) Traditional Supervisory Control
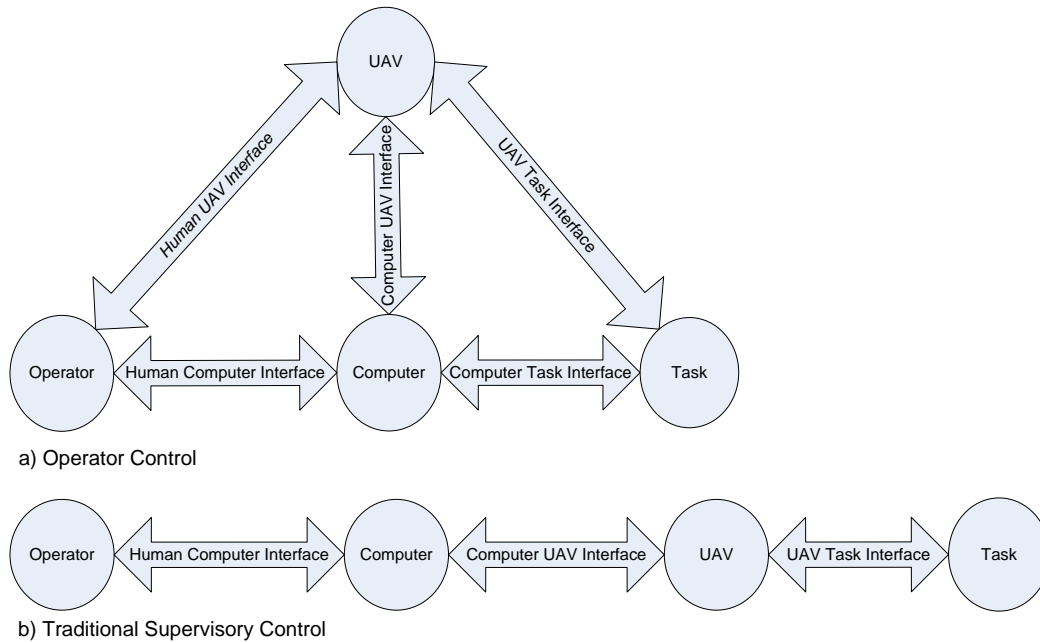
Figure 2.13 Evolution of supervisory control

As Figure 2.13.a illustrates, the operator performs a task with computer assistance. Since the levels of automation can vary the task may be performed by the operator, the computer or combination of the two. The Human UAV interface is the mental representation the operator forms of the UAV's situation, *i.e.*, position, acceleration, etc. How well the operator can form this mental picture may determine how well the operator projects into the UAVs situation. There is no direct communication between the UAV and the operator. The human directs the computer as to which global and intermediate goals need to be performed; in turn the computer directs the UAV to perform a task (fly left, land, search, etc.); and finally, the UAV performs the task. The UAV keeps an internal representation of its neighborhood and progress towards its task achievement. The UAV reports on task progress by sending data back to the computer.

The computer uses data incoming from the UAV and internal data structure representations to update the operator through the Human Computer Interface.

The traditional supervisory role shown in Figure 2.13.b. illustrates how the supervisor/operator no longer has direct contact with the task. The computer directs the UAV to perform the task while the supervisor only corrects any perceived variance from the desired outcome. The levels of automation may still in theory vary, but in a practical sense, supervisory control is useful when the computerized system has at least a moderate level of sophistication.

2.2.5.1 Human Supervisor Tasks

With these automation guidelines, some of the human supervisory roles for multiple UAV control might include: 1) mission assignment, 2) subgoal assignment, 3) resource allocation, and 4) mission algorithm assignment. For instance, the mission's overall goal may be to identify survivors after a hurricane, but the individual UAVs may be assigned to search different map areas. Furthermore, the supervisor may direct each UAV to search differently depending on the priorities and exigent circumstances.

The supervisor must trust the UAVs to have a minimum level of autonomy, for example, they must be able to continue performing assigned tasks without intervention. While important decisions are necessarily reserved for the supervisor, this does not mean that the supervisor is not intimately involved in the perception of the sensed data, but rather that the UAV agent must also be highly automated in perceiving whether a sensed blob is a target and what that target's preliminary priority should be. If the sensed target

is easily identified as a target by the UAV, why not utilize that information? The supervisor cannot track all sensors in a multiple UAV group simultaneously to identify targets and thus he requires help from the UAV in target identification. However, not all blobs are going to be easily identified by the UAV and having a human's pattern recognition ability complements the UAV's ability.

Along the three axes, the most difficult axis to model is determining the appropriate level of automation. As time progresses within a dynamic system, the operator's and UAV's priorities and goals are going to change. For example, the internal priorities of targets may change over time in a SEAD mission. Suppose a UAV identifies a low priority stationary target during the initial search. The location is marked using virtual pheromones and the supervisor's global view is updated. The decision is made (by the supervisor or UAV) to continue searching in hopes of finding a higher priority target. As the mission progresses and fuel levels start to get low, suppose no new target are identified. Then the UAV raises the priority of the already identified target and is scheduled for attack. On the other hand, suppose a new higher priority target is found. This higher priority target is attacked, while the supervisor has the option of assigning a different nearby UAV to attacking the lower priority target. In this way, the UAV resource is efficiently and optimally used.

2.3 *Summary*

This research overview presented several theories that individually are incomplete as design tools. The theories' creators (SA, NDM, etc.) purport that their specific theory is the complete design tool, but in fact, they should be used as tools within a toolkit.

According to Klein (2009), Situation Awareness fails as a theory because it does not use the historical information when analyzing the designed display, but is in fact a measurement tool that gauges the operator's understanding of the present and future states without formally addressing past circumstances. The fact that the operator may use their knowledge of past events (experience) to predict future possible events is not explicitly addressed by SA. On the other hand, Naturalistic Decision Making encourages the context sensitive design of user interfaces, without really addressing how to assess their validity, nor does it present a specific systematic design framework. NDM does support the formal interview process before and after the design to ensure that the stakeholder's needs are met but the design is *ad hoc*. Missing from both of these tools are the psychological and physiological tools needed to design the display, which is addressed by visual thinking theory. Visual thinking emphasizes designing the display to exploit basic visual pre-attentive cues, giving the operator the clues needed to make the correct decision. And finally, semiotic analysis is the linkage between the domain culture (context) and the display elements – how and why a display is used within the specific domain.

Hence, each of these tools is purported by their supporters to be the complete "design" package, when in fact they perform admirably as tools within a larger toolkit. The first phase of this experiment is to develop a systematic approach that can be used by field practitioners to design human centered, cognitively sensitive displays. The second phase incorporated a more robust display that allowed the operator to actively monitor the system, using historical system data to make predictions about near future states. The

third phase integrated the developed displays into a simulator test bed and was observed

by current UAV operators.

3 Semiosis/Semiotic Analysis of UAV Command and Control User Interface

As mentioned in Chapter 2, Semiosis is the study of signs; including the cultural influences on the interpretation of signs. What follows is an examination of semiotic analysis applied to the different displays used in the command and control user interfaces.

3.1 *Semiotic Analysis of UAV command and control system*

Semiotic analyses applied to the domain of UAV command and control concludes the following observations.

(1) Not all data available to the controller/operator is graphic. Reams of paper are generated daily to inform the operator of weather, intelligence, target destinations, etc. Some of this information is presented as images (weather patterns overlaid on a map, target location, etc.), while some is listed in tabular text form (available UAVs, target importance, target location, mission goal, etc.).

(2) Information is processed through multiple data pipelines, informally known as stovepipes. This limitation is in part forced on the system by onerous US government regulations. Whenever a secure system is accepted by the US government, it must pass a series of difficult tests. This testing can take over a year! So when an accepted system needs to add a new data feed, the entire system needs to be recertified. One way around this is to have the data

stovepipe certified separately from the rest of the system. This requires the addition of separate displays for the new video displays. (At one time, there were five screens displaying separate, disjoint views of the mission.) Or, as is the case with some data stovepipes in the UAV control domain, hardcopy text table printouts that the operator reads/skims prior to and during mission runs.

(3) Not all information could be quickly accessed from the operator's chair. Some information is posted on a clipboard located away from the control station. On shift changes, operators tended to read this information prior to replacing current operators occupying the chair. Changes to the environment (weather, social, mission, etc.) may be represented on one screen, and then mentally composited on other screens by the operator.

(4) Most monitoring displays mimic existing paper/pencil tables or existing gauges. In general, the displays do not take advantage of inherent human cognitive perception abilities. This tends to make decisions more difficult to arrive at and takes longer to execute. Most of the time, this time delay is not crucial, but there does exist the possibility of a catastrophic result if the decision is delayed or wrong.

(5) Since there is more than one person controlling the UAV, each person has his own responsibility. It follows that there are times that the information needed by one member is known by another or must be retrieved by another.

(6) Not all of the mission parameters are available to the operator prior to the UAV mission launch. Since we are in a highly dynamic environment, missions, goals and targets can change quickly and unexpectedly. Intelligence

operators frequently hi-jack the UAV mission to accomplish new, high priority goals. Forces in contact with enemy forces drastically alter priorities. Informal and formal socio-political environments can alter UAV path parameters profoundly without having an associated "data pipeline". Furthermore, adverse weather conditions can affect priorities on the mission list.

One of the overarching themes of the semiotic analysis is the disjointedness of the data/information pipelines. In part, this research presents a flexible display that can have multiple data pipelines that can be visually organized as a static, weighted, figures of merit decision aide or a dynamic, weighted health monitoring aide. Furthermore, the number of axis available to view by the supervisor is flexible (range: 6-15), allowing the supervisor to dynamically remove any inconsequential axis to further aide decision making.

3.2 *Semiotic Analysis of Alternative Interfaces*

Next, we perform a semiotic analysis the different displays that could be considered as decision aids within the multi-UAV command and control domain and look at their strengths and weaknesses. First, we examine the original data pipeline – text tables. This is followed by the graphical displays: bar chart for the static decision aid display and analog gauge for the dynamic monitoring display.

3.2.1 Text Display

As mentioned previously, the default data access to the operator for UAV missions is the daily operations report, typically posted by the door, and more importantly, not digital. These pages of text include daily weather reports, expected tactical and strategic conditions, expected troop and weaponry locations (both friend and foe), no-fly-zone information, etc. What it does not contain is information concerning specific mission requirements, i.e., it does not contain the route the UAV is expected to travel.

These original data tables were typically presented in generic tabular form on several pages – the operator must be trained on where the important data is located. Within this existing environment, there exist no cognitive cues to dynamically assist the operator. The tables are plain black and white, without much formatting. Because of this generic format, the learning curve is steep, and long experience of the tabular format must be endured to become proficient. Furthermore, the data tables can be presented on a hardcopy daily report or (a more recent innovation) through a separate screen monitor. These two sources of data were not necessarily coordinated, nor consistent. The operator's experience is used to analyze and decipher any irregularities. Furthermore, digital tables may not be updated regularly, so the operator must be aware of possible time lags.

This requirement of a learning curve to achieve reasonable mastery must make one conclude that the text table in any of its forms is an iconic symbol. In fact, tables as a generic information conduit are an iconic symbol system that is learned as a child in

elementary school mathematics. One speaks of an iconic symbol **system** because the table as an entity has many conventions involving row and column manipulations (sorting, location, etc.) to speed data access and interpretation. Furthermore, although the existing tables do not use visual cues, there exist some standard visual cues that can be incorporated into the tables to aid emphasis of important data.

Static Text Table Display

Relative importance or weight of the 7 dimensions

Dimensions of route merit; all describe factors contributing to mission success

Three routes with their corresponding dimension scores

Absolute Criterion needed to be a suitable route

Red text Indicates FOM failure

Direction of dimension, +=higher is better, -=lower is better

Weighting	Figure of Merit	Route 1	Route 2	Route 3	Criterion	Better
40	probabilityOfSurvival	0.5	0.3	0.2	0.2	+
15	samShots	3	4	9	10	-
15	aiLaunch	5	3	2	8	-
15	aaaExposure	200	500	600	300	-
9	netNumTrack	4	2	11	15	-
3	samTrackingRadarPd	0.2	0.5	0.7	0.6	-
3	ewRadarPd	0.6	0.4	0.9	0.8	-

Figure 3.1 Static one-threshold text table display

For this research, the text tables were adapted to show upper and lower threshold through color cueing. In the first experiment (Figure 3.1), the text tables were static with a single threshold indicated by a **red** number, while the second experiment (Figure 3.2) had dynamic text tables that were updated every second and had a low threshold and high

threshold identified by a **blue** and **red** number, respectively. These choices of color were chosen to reflect the western culture cues of cold **(blue)** and hot **(red)**. Although not displayed in this figure, the dynamic displays had an option of displaying the low and high thresholds in the lower left and right corners. This was useful in predicting the possibility of the UAV going out of tolerance in the near future (SA3).
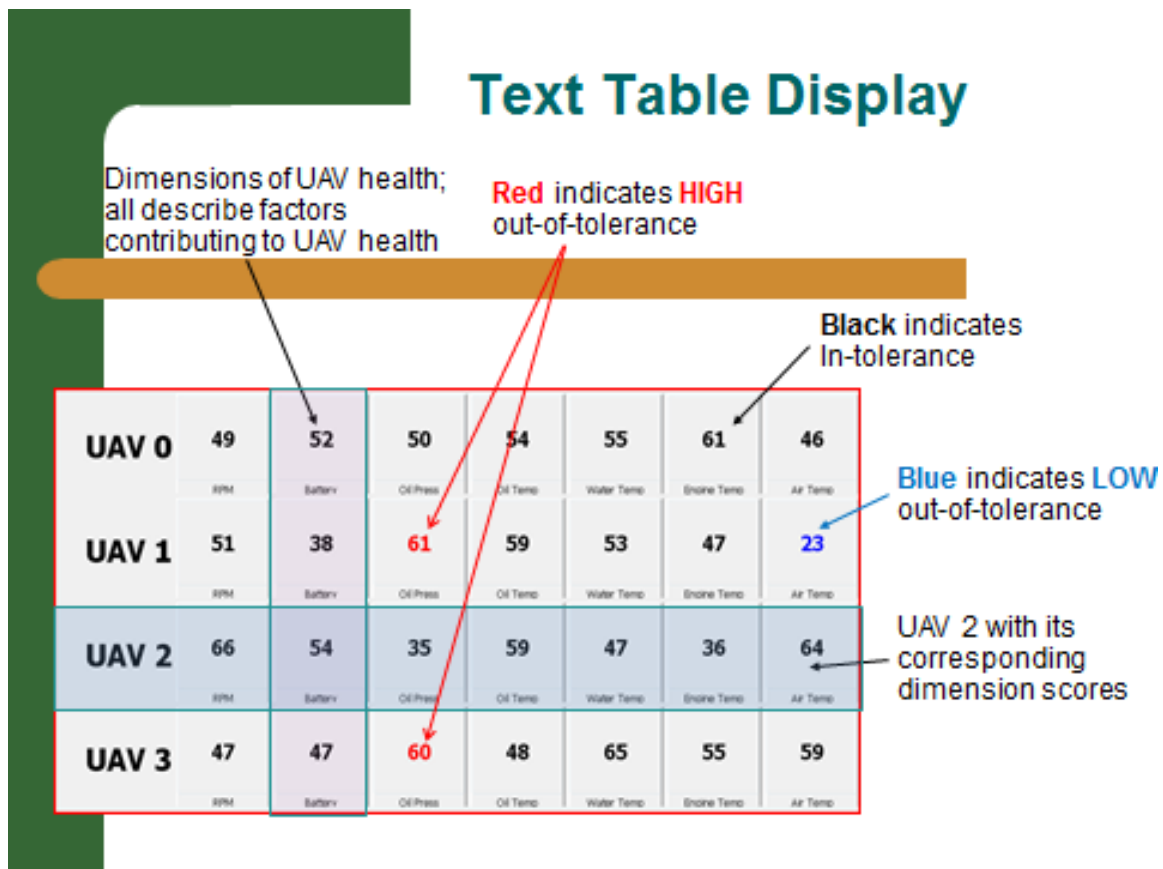
Figure 3.2 Text table display for health monitoring system

3.2.2 Bar Chart – across and between axis

The bar chart is one of the subjects of middle school mathematics, and thus is also an iconic sign system. Traditionally, the bar chart presents a relative scaling along one

axis as a simple graph or as nested bar charts to examine multiple axes. In the UAV domain, it can be presented to the operator in three versions: 1) across the Figure of Merit (Anti-Aircraft batteries, distance, probability of success, etc.); 2) across the UAV (UAV1, UAV2, etc.); or 3) a nested hybrid (UAV with nested FoM or FoM with Nested UAV). The single axis charts are simpler to interpret, but have limited usefulness – if your question about a FoM is presented via a UAV mapped bar chart, then the operator may struggle to find an interpretation. The nested bars are difficult to interpret on a monitor due to lack of screen space.
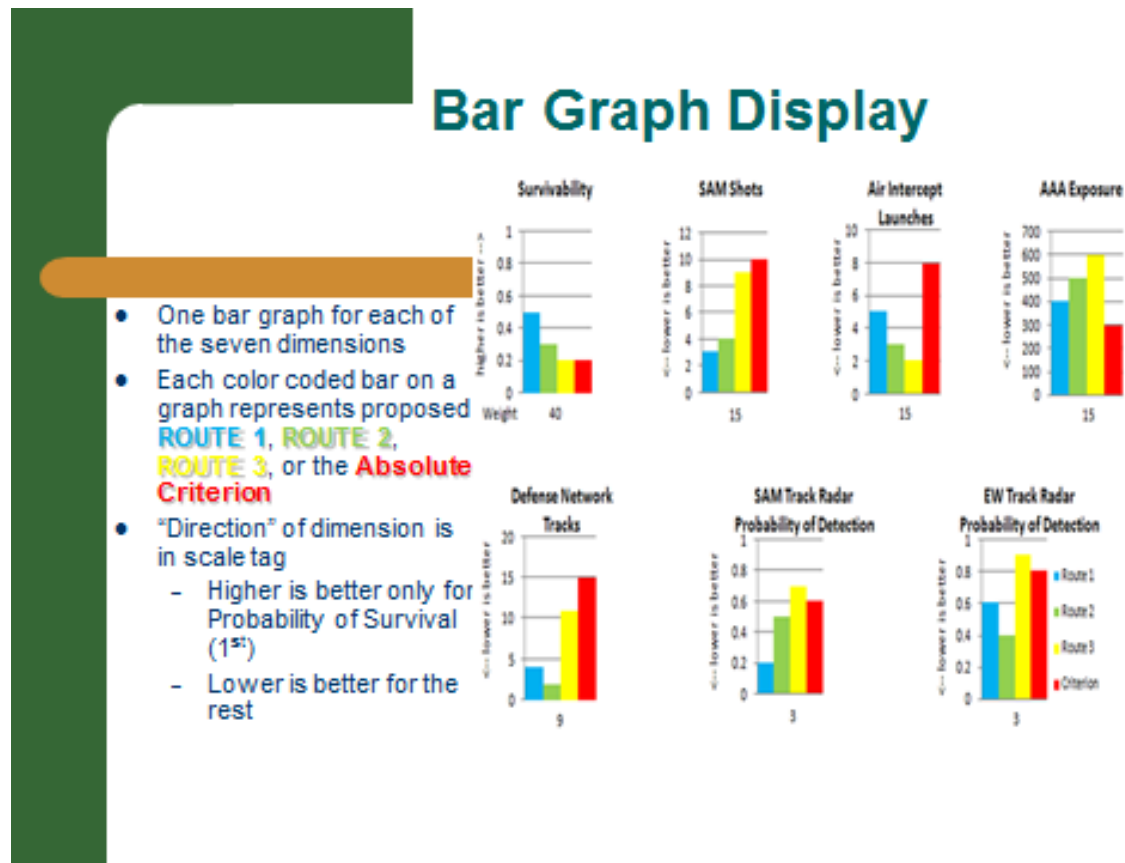
Figure 3.3 Bar graph display as a decision aid

The bar graph presented in this research (Figure 3.3) chose to present across the FoMs, and thus were handicapped for any questions across the UAV dimensions. Along the left side, the label indicated whether the information presented was better with a large value (higher survival rate) or a small value (fewer SAM sites, Air Interceptors, etc.).

3.2.3 Analog Gauge

The analog gauge implementation derives from the analog gauge currently supplied in automobiles and aircraft, and is thus a pervasive iconic sign system. The normalization of the data sets to range between [0-100] allowed for standardized gauges that arced 180° through this range. Furthermore, each gauge had their high and low thresholds identified by **blue** and **red** sub-arcs.
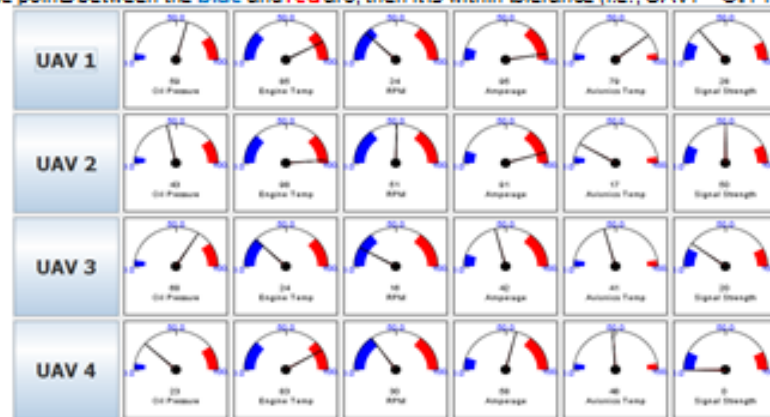
Figure 3.4 Analog gauge display for health monitoring system

This GUI was implemented as a naïve approach a graphical display for multiple UAVs – it is easy to implement (existing software libraries have gauge widgets) and it has a direct real-life analogue to the gauges currently found in aircraft.

3.2.4 Visual Thinking Interface Design

During development of the Multiple UAV Agency (MAGE), designed to integrate net centric information to facilitate control of multiple UAVs, one of the chosen features for the MAGE software was to provide automated mission planning to reduce operator workload. Towards this end, SYTRONICS licensed from Operations Research Concepts Applied (ORCA) its OPUS mission planning software library. OPUS was employed by MAGE to generate

multiple alternative mission routes from which the MAGE operator could choose to execute. In addition to the routes, the OPUS software also generates a series of "figures of merit" (FoM) to describe the characteristics of each alternative. These FoMs include more than a dozen dimensions, including the probability of surviving the mission, number of surface-to-air missile launches, minutes of exposure to anti-aircraft artillery, minutes exposure to search radars, minutes exposure to missile guidance radars, fuel consumption, to name a few. These measures are all numeric but have different measures, different minimums and maximums, and run in different directions (fewer SAM missile launches and higher probability of survival are both better).

The multiple scales, dimensions, and directions make it very difficult to integrate their meaning to get an overall idea of a route's overall merit. Our original representation of the FoMs was made in tables, similar to those shown in ORCA's own interface to the OPUS mission planning software library. However, because of the number and complexity of the FoMs the tables were not effective in supporting route alternative choices. Our government technical contacts asked that we develop a more effective representation to support choice of route alternatives.

We chose to approach the design problem from a cognitive systems engineering perspective, looking to theoretic psychology to determine how to build the new FoM display. One particularly promising cognitive theory was Arnheim (1969) description of what he called "Visual Thinking" (VT). Arnheim was trying to explain how artists, particularly painters, shape their works cognitively. He theorized that human cognition evolved from sensory information processing, particularly visual information processing, and shared communality with the perceptual processes. Although Arnheim never considered the implications of this theory to displays and controls, others did. McKim (1972) tried to translate VT's hypothesis to the

information design to facilitate problem solving. Using McKim as a point of departure, our design effort prototyped displays based on perceptual-cognitive principles (Figure 3.5).
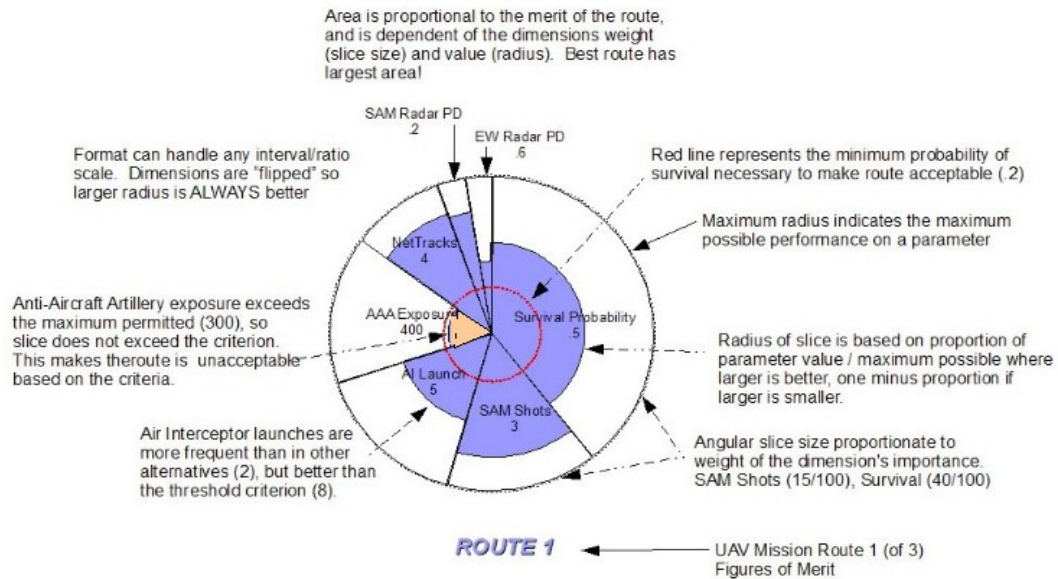
Figure 3.5 Early drawing of a multi-dimensional, multiple scaled decision support display. Major difference is the raw data display only was visible by mouse roll-over of the pie slice.

The acute perceptual skill we based on display design was form recognition and in particular, the ability to judge the relative area of forms (Cleveland and McGill 1984). Experimenting with a large number of different representations of the FoM data, we derived a "pie" representation that employed expensive rescaling and standardization of the different route effectiveness measures. The process first called for first determining the relative weight of each FoM to the final decision. These weights determined the angular subtense of each measure's slice of the pie.

Second, a minimum and maximum value was identified for each measure. This process was limited by identifying values beyond which there was no significant difference to the decision maker to keep the dynamic range within manageable values. Also determined was a

threshold value (high or low) which defined whether the FoM made the mission alternative acceptable.

Third, reciprocals were taken of those scales which smaller values were better. This "flipped" these scales so bigger always meant the preferred condition. Finally, the values falling below the threshold were linearly scaled between their smallest (worst) values and fixed threshold radius. This identified route FoMs that did not make criteria and their area was shaded red to flag their failure to achieve the stated tolerance. A minimum radius was applied so there was sufficient area to signal the user of the violated tolerance. Values above the threshold were scaled between the threshold and the maximum diameter of the pie display. Hence, these pie slice radii represented acceptable FoM values and their areas colored blue.

The resulting circular figure we called a "Sprocket" because it resembled the tooth embellished wheel that drove a chain, or in this case, cognitive understanding. It implements the VT paradigm because human vision is adept at area judgments and the Sprocket is designed so that its area represents the overall merit of the alternative route. The user at a glance can compare the area subtended by each alternative's FoMs to determine the best route. It accomplishes graphically what a weighed sum of rescaled values does mathematically. However, it clearly displays which dimensions contribute to the overall worth of the route and which dimensions exceed or fail to exceed their threshold requirements. Further, if the decision analyst wishes to alter his or her weightings from the ones used to render the display, it is easy to visualize the changes and use them to alter the decision.

4 Research Framework

This research used the cognitive systems in context framework developed by McNeese, Bautsch *et al*. (1999) to design and evaluate the user interfaces for multi-UAV supervisory control (Figure 4.1). This general framework "provides a continuing specification of boundaries around a work domain", the research domain being Multi-UAV supervisory control. McNeese's research framework is a guide for research experiments within a given context. A researcher must, implicitly or explicitly, begin with a set of goals. Based on these goals, further "decisions regarding the experimental world, knowledge acquisition methods, representational schemes and evaluative procedures" must be made. With each decision made, the researcher further limits the research plan, giving the researcher more awareness of what needs to be done, and specifying the level of detail required to complete the research. In essence, these constraints narrow the choices the researcher makes, excluding choices that do not make sense given the previous decisions.
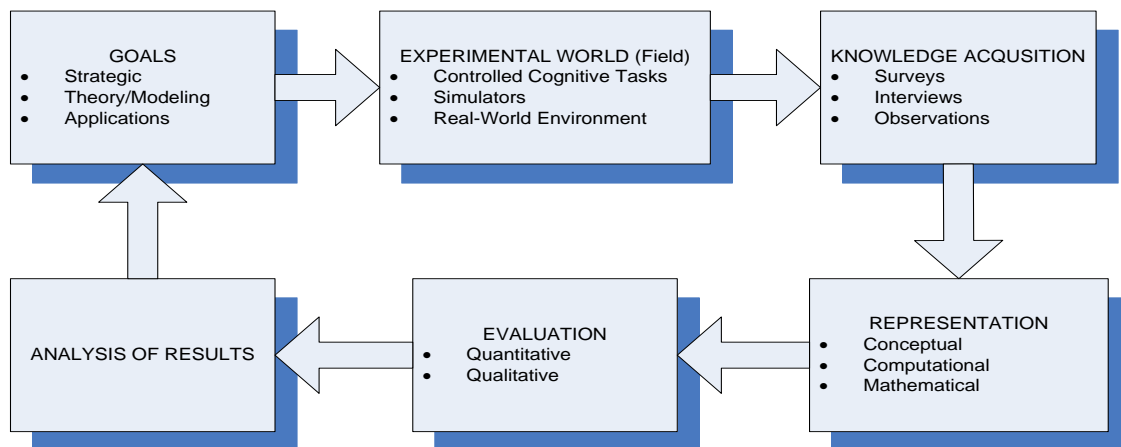
Figure 4.1   Components in the study of cognitive systems in context [adapted from (McNeese, Bautsch *et al*. 1999)]

As an example, this research's goal is to evaluate visual thinking widgets as a decision aid and monitoring displays in the multi-UAV command and control domain. The expansion of Figure 4.1 showing specific goals for each experiment in this study is illustrated in Figure 4.2. The figure explicitly illustrates the three experimental blocks (named Exp.1, Exp. 2, and Exp. 3) and the decisions made for each. For each block in the figure, a decision had to be made with regards to the individual experiments.
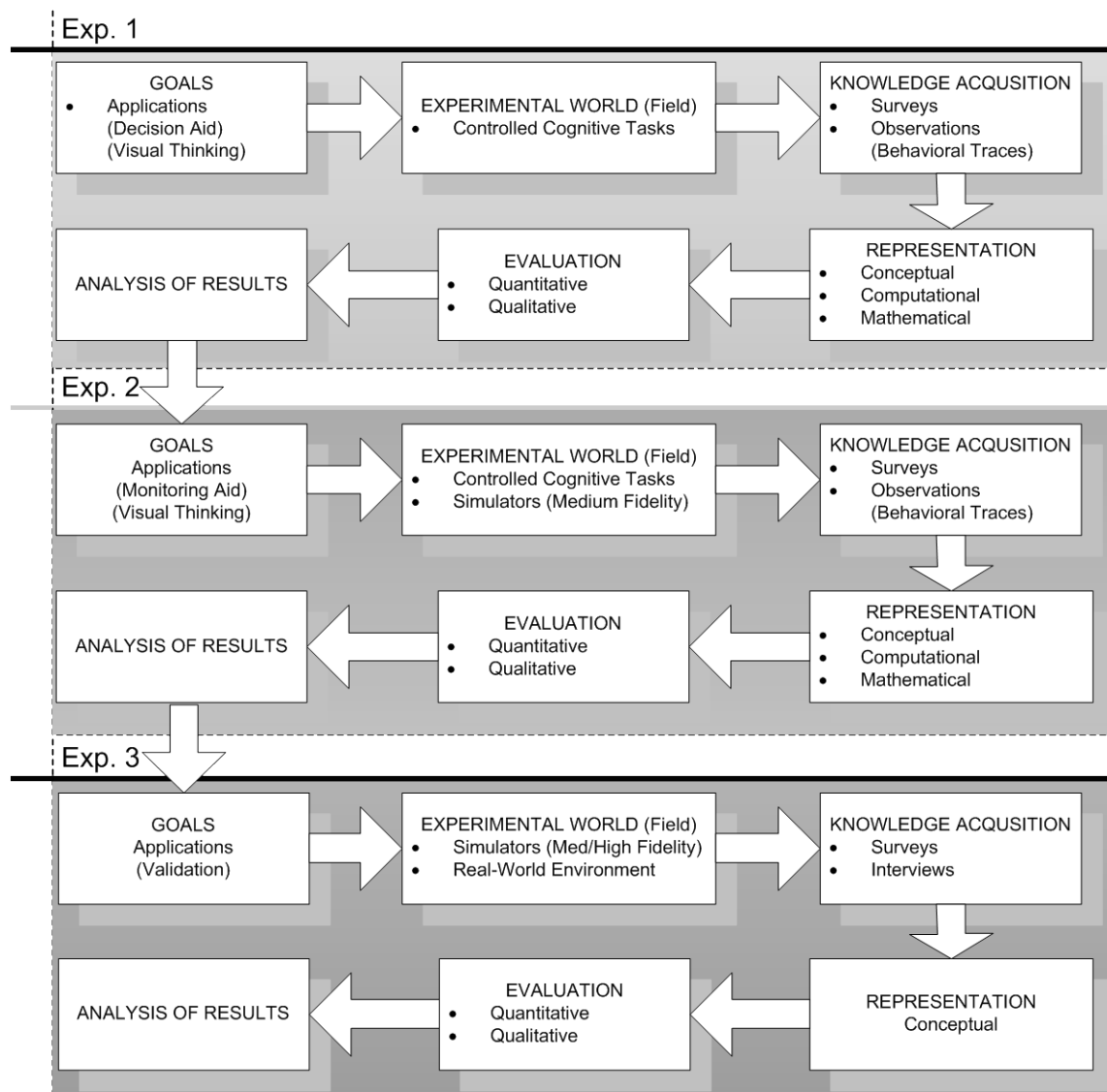
Figure 4.2 Framework of cognitive system expanded to this research

However, Figure 4.2 is visually overwhelming and rather complicated, so each experiment will be discussed in more detail with a breakout of the figure above to emphasize the decisions made along the way. Along with this extraction, the questions each experiment was going to answer are listed as hypothesis.

4.1 *Experiment 1: Decision Support Research Framework Details*

The first experiment's goal is to use examine the visual thinking widget's viability as a decision aid, informally called a visual thinking sprocket (or VTS) in honor of the cartoon "The Jetsons" and "Spacely's Sprockets" factory.   The visual sprocket is compared to other standard text/graphical decision aid displays.   The McNeese framework for Experiment 1 (Figure 4.3) explicitly illustrates the decisions made for this experiment.
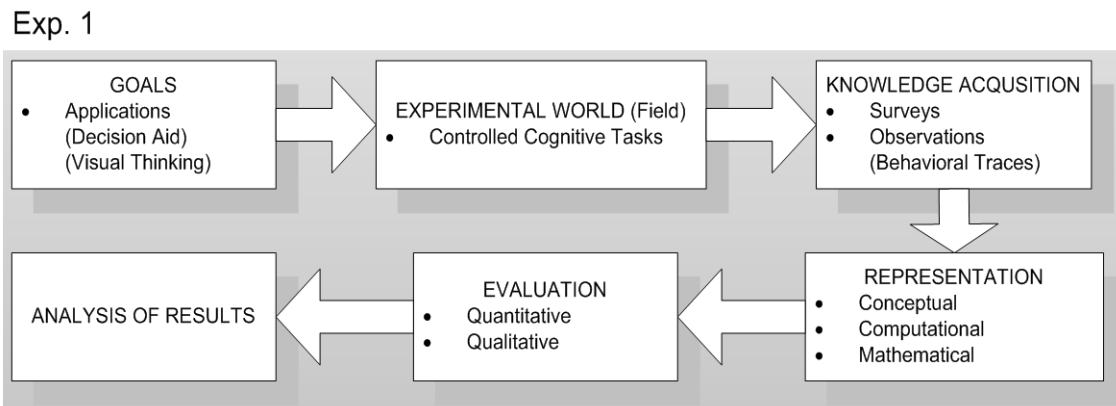
Figure 4.3 Framework for Experiment 1

From the first block, one sees that the goals of this experiment were to design and evaluate a visual thinking based user display as a decision-aiding tool for Multi-UAV command and control mission assignments by UAV supervisors. After discussion with

stakeholders within the domain, it was determined to use the visual thinking sprocket as the decision aiding display.

As shown in block 2, from this decision and the fact that UAV operators are not readily available for experiment subjects, it followed that a very restrictive laboratory setting was appropriate for the experiment. This laboratory environment allowed the research to utilize college students as the subjects in the experiment. As an added benefit, using college students with no prior knowledge of UAV operations removed any potential learning bias that a UAV operator subject might have with the existing system.

Block 3 of the framework included questions generated during the experiment addressing different levels of situation awareness (SA) and questionnaires after each block (Appendix B). In a cognitive system, the actual system functionality, rather than the theoretical or ideal function is important (Hollnagel and Woods 1999). Thus, any new proposed system must be empirically evaluated. This empirical evaluation used SA directed questions to instrument the display, posing questions to the subject. The question format and applicable SA level is presented in Table 5.1.

Table 4.1 Experiment 1 (Static Displays) questions

Situation Awareness Level	Question	Possible Answers	Abbreviation
Level 2 SA	Rank order the routes [Best to worst]	1-2-3 1-3-2 2-1-3 2-3-1 3-1-2 3-2-1	Rank_Order
Level 1 SA	Do any of the routes meet the all minimum criteria?	Yes No	Minimum_Criteria
Level 3 SA	Which route is Best if Dimension X is dropped? (Where X was chosen from among the 4 top weighted available dimensions)	1 2 3	Drop_Dimension
Level 1 SA Level 2 SA	Which route has Best Dimension Y? (Where Y was chosen from all available dimensions)	1 2 3	Best_Dimension

To judge the effectiveness of the new sprocket, data were collected both automatically (behavior traces) via the computer and manually after each block (each subject was asked to subjectively evaluate each display type via a questionnaire after each block). After all three blocks of the experiment were completed; a final questionnaire asked the subjects to rate their subjective preferences of each display type. The data automatically collected objectively reflects the speed and accuracy of the decision aid's usefulness. The data collected by the computer was statistically analyzed using SAS 9.2 to examine the potential of the visual thinking widget as a viable decision aid (Blocks 5 & 6).
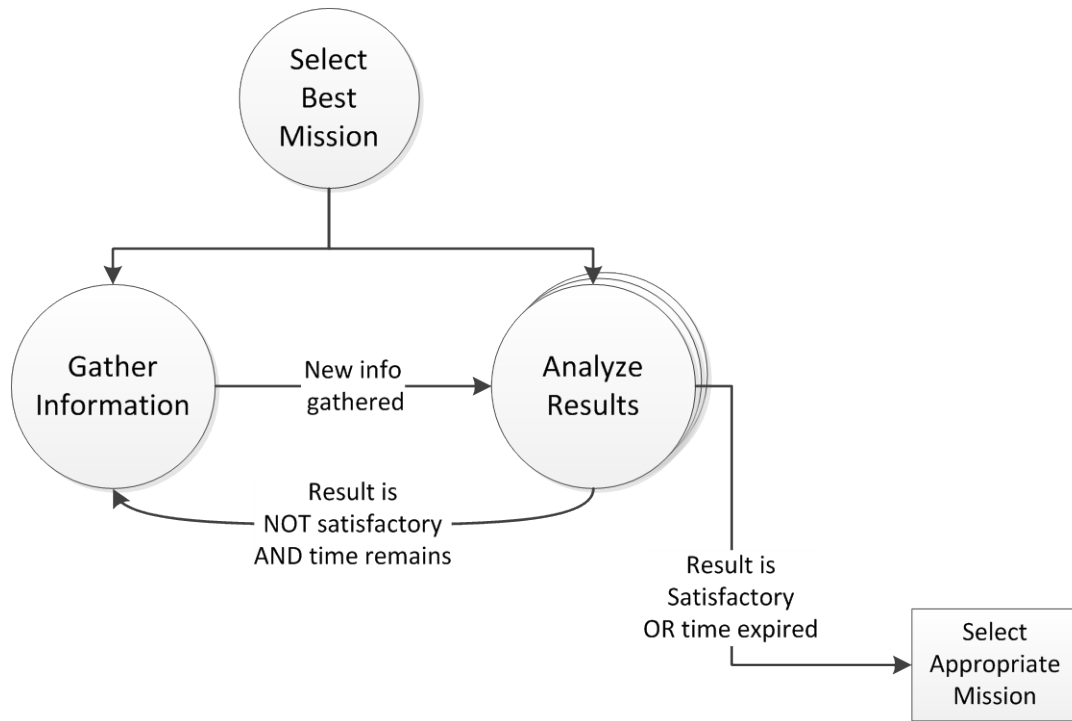
Figure 4.4 Overview of operator decision process

Modeling the operator's decision process for selecting the best mission route among those presented (Block 4); Figure 4.4 presents one very high-level representation of the decision process. As mentioned earlier, observations of current Ground Control Station (GCS) standard operating procedures show that the data the operators need to help make these decisions among alternative UAV mission paths is, in part, presented in text tables. These existing tables have no aids (highlighting or hints) to help the operator to speed the decision. Some of the data is available as hardcopy and while some is accessed through multiple computer displays. Furthermore, there is some overlap on the data available through the different data stovepipes; only operator experience dictates

which data stovepipe to access.[9] However, not all information necessary to make these decisions are available, so the operator must mentally construct or interpolate the necessary information. Rather than use the existing data tables, considered rather bohemian when compared with the sprocket display, the text tables were "updated" to automated displays that have visual cues to draw the operator's attention. In other words, the electronic tables had color-coding to emphasize any out of tolerance Figure of Merit.

Finally, it should be noted that the UAV operator does not receive a list of alternative mission routes as a rule, but rather each alternate mission route is generated and evaluated much like Klein's Naturalistic Decision Making (1993) satisficing principle. If the new route generated satisfied the minimum requirements, then select it and do not search for better solutions. If the new route failed to satisfy the minimum requirements (FoMs), try to identify reasons for failure and generate new alternatives that address some or all of the shortcomings. The method of generating these new alternatives is dependent on the operator, but generally appears to be the modification of previously successful mission routes that were applicable to similar situations. Once again, operator experience widens the pool of alternative routes generated.

On a visit in the summer of 2007 to a UAV command station[10], it became obvious that the necessary information gathered for the operator to select among alternative

---

[9] An information stovepipe is a data pipeline that largely restricts the flow of information to vertical communication and inhibits or prevents cross communication.

[10] The trailer came with four UAV operator control stations, each control station had three monitors (with room for 6) and a single notebook/clipboard hanging by the door with the daily intelligence report.

mission routes began when the operator walked into the trailer before mission handoff. The operator scanned the daily intelligence report hanging by the entrance. This took about fifteen minutes, and supplied the current mission objectives, applicable routing information (waypoints for the next 2-3 legs of the mission), weather forecast, Human Intelligence (FFN), etc. Following the intelligence report scan the new operator spoke with the current operator staffing the controls to get a flavor for the current mission status. At this same time, the new operator would be scanning the Mission Monitor, the Satellite/weather Monitor and the Decision Support Monitor. Each monitor had different information that operator might need, but no monitor had all the information that was necessary. This debriefing also took about fifteen minutes, after which the new operator changed seats with the old operator. The old operator might remain to observe the mission for a few minutes to verify all pertinent mission information had been passed on to the new operator. The old operator was then able to leave the trailer.

The reason for having many different places to get information is in part due to the government's burdensome acquisition requirements. To validate any new user interface requires a large investment of time, money and personnel – it takes between one and two years to validate and accept any new data system. If the data is to be added to an existing interface, the old interface might become invalidated until the new data is added. Because of the inordinate amount of time to validate this new data, when new data is deemed to be an important enough to add to the incoming data flow, a new separate data stovepipe is created. Each of the three existing monitors (Mission, Satellite/weather, and Decision Support) took more than a year to pass inspection. Adding another piece of information to any of the UIs would require revalidating that entire subsystem. Rather

than combine all user interfaces into a single complete data monitor, additional monitors are added; thus speeding up acquisition and lowering costs.

Returning to the UAV mission script, each UAV's complete mission may be secret and often it is on a need to know basis, and since the UAV could be in the air over multiple operator shifts, the operators frequently saw a partial release of the mission details based on time and the current location of the UAV. Furthermore, the UAV's mission may evolve over the course of the flight as targets that are more important and sub-missions rise to the surface of planning sessions. Generally, the mission planning was performed away from the GCS. (During an informal discussion session, one operator described flying UAVs as "herding cattle from point A to point B".)

Frequently, only the next waypoint or two in the mission were supplied to the operator, not the whole route. If an operator completed a UAV's sub-mission, then the mission "script" was modified or updated and the operator was then given the next waypoint of the current mission. In the case of a drastic mission modification, operators might be requested to assist in rerouting the UAV through the new mission. This new mission might include routing to a single new waypoint or a complete route with multiple waypoints that eventually returns the UAV to base. The operator must use their own experience and current environmental information to form a viable route that satisfies any mission requirements.

As a hypothetical example, imagine you are an operator flying a UAV on a typical surveillance mission that requires the recognizance of a suspected terrorist cell meeting. The supplied routing for this mission was given by the Information/Intelligence

Officer to ensure maximum stealth (hovering downwind so propeller noise is not propagated to the target and not easily identified by the naked eye) and UAV survivability (outside the range of known AAA/radar sites). Another way of saying this is that the top priorities of this mission from the commanders viewpoint is for the UAV to have a high level of survivability and stealth (not being seen or heard), while at the same time be able to fulfill its mission objectives to gather useful intelligence (UAV may have to adjust location to take pictures). In this hypothetical situation, new information comes through channels stating that ambushed troops "nearby" are presently under fire – a UAV is needed to support the troops as their "eye in the sky" and to aid ground support response planning. One needs to gather information to determine which UAV(s) should/can be reassigned in this high priority mission (usually distance and speed are the dominating parameters). The UAV that is closest to the event and has a sufficiently low priority mission can be reassigned – so the UAV is reassigned to this "support the troops" mission. Now the highest priorities for this new mission might be speed and survivability en route; however, after arriving the mission priorities might require stealth again. Where the priority on a simple surveillance mission might require survivability and stealth over speed, this new support mission might initially require speed over survivability and stealth. After all, human lives are at stake and time is crucial. Furthermore, if the UAV is armed, then the UAV provides some valuable missile support to the troops on the ground. If the UAV is not armed, then it might supply temporary visual support while waiting for an armed aerial response to arrive.

The previous example illustrates the extremely complicated decision process that the operator may perform while responding to a change of UAV missions. This decision

process can be modeled by an Operator Function Model (OFM) (McCormick 1976; Miller 1985; Mitchell and Miller 1986; Narayanan, Ruff *et al*. 2000).

4.1.1 Operator Function Model

An OFM is a mathematical construct finite state automata network of arcs and nodes that represents the cognitive processing of a process. If represents the mental decomposition an operator might use to breakdown a complex system into simpler parts and any associated actions necessary to guarantee an acceptable overall system performance. Other researchers have used OFMs to successfully model, design and control user-interfaces and supervisory control systems. OFMs represent knowledge representation, information flow and decision-making in complex systems (Narayanan, Ruff *et al*. 2000).

The previous hypothetical example also alludes to the mental effort required by the operator to determine a new UAV mission route. How does the UAV operator determine the Probability of Survival (PoS)? What type of data needs to be gathered? Where is it located? Is the data out of date? It should be obvious that the data is incomplete and out of date (If it were complete, then the troops under attack would have already had air support.). So the operator must gather information to make a reasonable decision (Figure 4.5). From the figure, one can see that information is essentially pulled from four sources: mission, weather/satellite, decision support monitors and an "annotated" printout of current and recent satellite images and intelligence reports. (The standard printout is very poorly annotated as black and white dense tables with labels, but little else to identify important datum. No data or trend analysis is available.)

The mission monitor supplies a map of the terrain along with an overlay of the active mission route. It also contains icons of relevant mission objectives and landmarks, such as targets, friend/foe/neutral, AAA sites, EW radar sites, etc. The symbology assists the operator, but the information is frequently out of date and incorrect. The location of enemy troops can be especially problematic in a non-traditional war, since the enemy moves fast and can be quickly camouflaged as civilians.

The weather/satellite monitor can give current weather conditions as long as the operator is experienced enough to interpret the map. However, this monitor does not supply a forecast capability. To the public, the weather forecast can be easily obtained through a website (http://www.weather.com), which has international maps that supply hourly forecasts. However, the UAV operator does not have access to a "web enabled" system inside the ground control station. An external source supplied the weather forecasts (daily intelligence logs or a dedicated system as side effects of the information stovepipes).
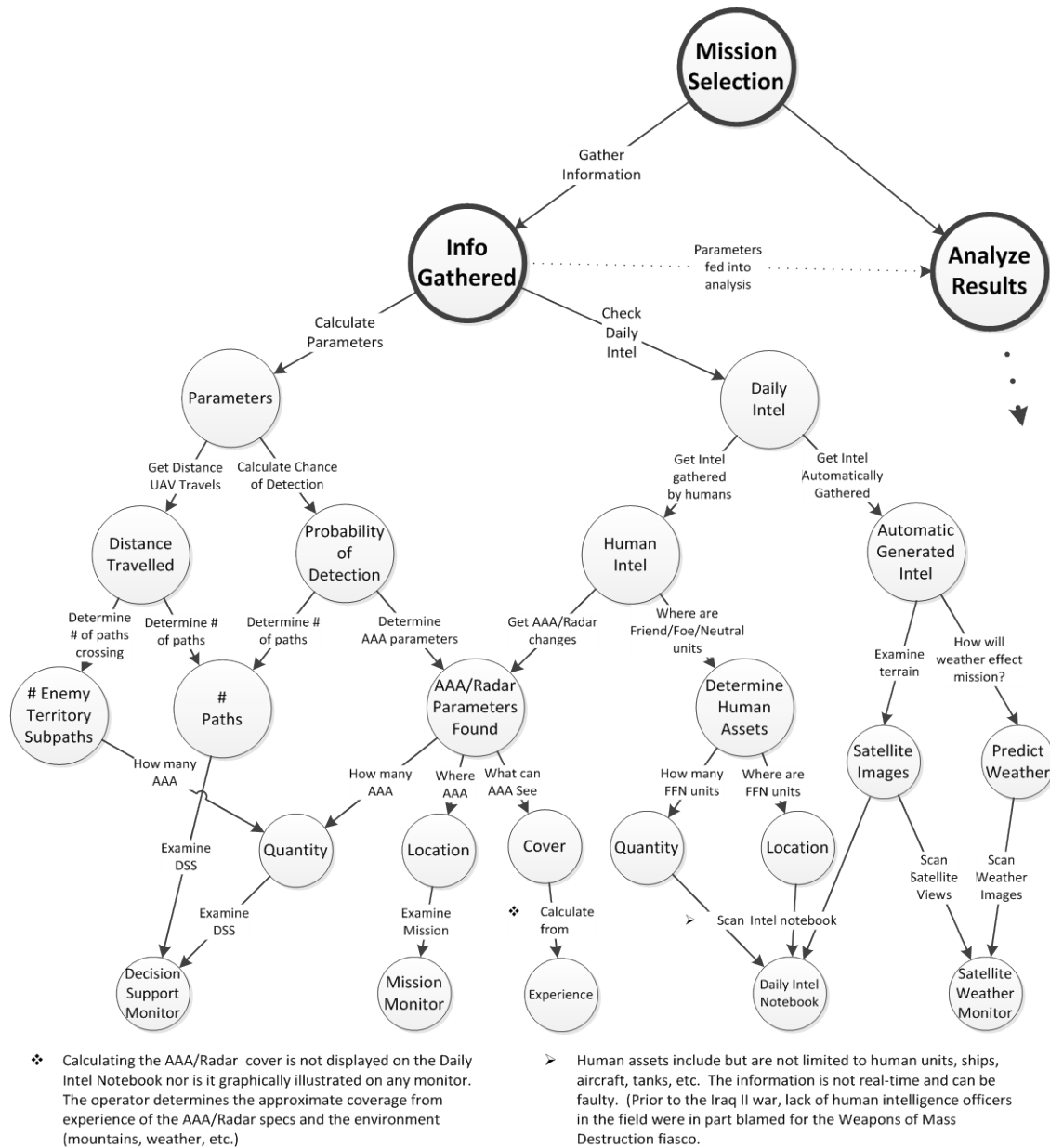
Figure 4.5 Operator Function Model representing the information gathering stage of the decision to select the best mission among alternatives.

Suggested mission alternatives can be partially compared on the decision support monitors. The alternative mission routes might be graphed on a map, but the map does not supply hints or Figures of Merit that would assist the operator in making any decision. The operator must attempt to mentally calculate path lengths, AAA and EW

radar footprints, etc. Furthermore, to compare the alternate mission routes to the current mission, the operator must mentally translate the current mission route on the Mission monitor onto the Decision Support monitor, since the current mission route was not displayed on the Decision Support monitor. In other words, one of the alternatives – "do nothing" – is not one of the choices on the Decision Support monitor. These are gaping failures of the current system.

Furthermore, the mission objectives described in the daily intelligence logs can be terse and incomplete. Not all of the UAVs tasks will necessarily be listed on the intelligence logs because of the "need to know" secrecy of the missions. After all, a mission could last up to 40 hours and the first operator might not need to know what the third operator is doing, nor does the third operator need to know what the first operator did. The current, recent past and near future events are all that is relevant to the operator.

Once the information is gathered, it must be analyzed by the operator as modeled in Figure 4.6. The information supplied on the data acquisition side of Figure 4.5 becomes the parameters to an optimization function side of Figure 4.6 in which the Probability of Survival (PoS) for the UAV is calculated. Now we examine how this PoS characteristic determined or computed. There are quite a few parameters that could make up the PoS, for instance, the weather, probability of detection by AAA sites, the number of AAA missiles available, airports, number of aircraft available for intercept at those airports, number of intercept aircraft in the air, etc. Some of these FoM are reported through real-time information pipelines, while others are updated at some longer discrete intervals (hourly, daily, etc.).
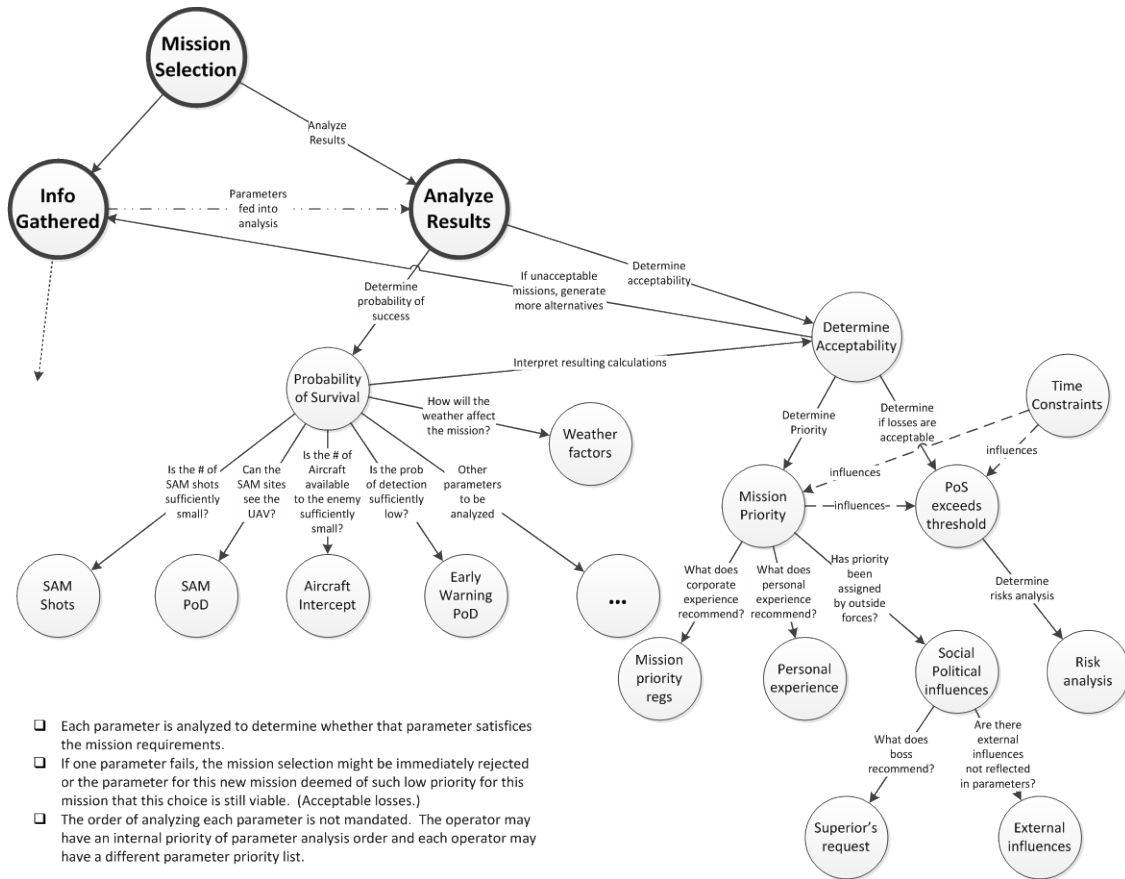
Figure 4.6 Analyze Results OFM

The operator selects the best mission route by calculating the PoS based on these supplied parameters. For example, given two mission routes, one route with six known SAM sites and the other route with only two, the second route is preferable, based solely on this one parameter. However, the decision is seldom based on one Figure of Merit (FoM) and sometimes the other FoMs suggest different alternatives. Since there are no documented procedures to calculate the PoS, the operator uses an internal and very personal algorithm to determine whether enough information has been acquired to make the decision. The OFM shown in Figure 4.6 is an example of what an operator might consider enough information. If it is not, then the operator attempts to gather more information to better aid in discriminating among the existing mission choices.

Alternatively, a new mission that better fits the accumulated data might be suggested by the gathered information.

Since the experiment examines an existing complex, dynamic decision support domain, UAV mission stakeholders were interviewed about what data was important to the operator when making mission decisions following the walkthrough of the UAV command station. As expected, there was a wide range of opinions in the number and subjective priority of FoMs to use. This led to designing the sprocket with the flexibility of displaying between six and fifteen FoMs with operator defined weighting. These discussions directly led to the eight FoM and associated subjective weightings used in the first experiment.

Based on the information derived from these interviews, a pattern of the decisions made by each operator emerged, namely an optimization problem. Each operator applied different subjective weightings to each of their parameters (FoMs) in the problem. For each of these FoMs, there existed an empirical threshold in which the alternate mission was scrubbed (more than 18 available SAMs, probability of success less than 25%, etc.). In these cases, other alternatives would have to be explored.

The use of each parameter required the operator to have experience to interpret the results quickly. Some of the apparent decision parameters for the operators could be best defined as composites of other simpler parameters (Probability of Success), while others were atomic (SAM sites). Some FoMs were better when bigger (probability of survival, etc.), while other FoMs were better when the parameter was smaller (probability of detection by X, number of AAA sites, time in battle zone, etc.). Furthermore, each

FoM had a different range of values. And finally, a given FoM may have a maximum range, but if exceeds a given threshold below that maximum, then the risk may be too great.

These disparate conceptual FoMs must be combined into one cohesive value representing the "goodness" of a mission route in order to differentiate among mission alternatives. As a first stab, they can be represented informally by the equation

$$f(x) = \sum \min(FoMs_{FOMs\,that\,need\,to\,be\,\min imized}) + \sum \max(FoMs_{FOMs\,that\,need\,to\,be\,\max imized}) + c,\text{ where}$$

c is the unknown, unaccounted for information that is not represented in the formal FoMs. Incorporating the threshold simply redefines the **min/max** functions to be the step

functions $\min_{Treshold}(x) = \begin{cases} \min(x) & if\ x \le threshold \\ \infty & otherwise \end{cases}$ and

$\max_{Threshold}(x) = \begin{cases} \max(x) & if\ x \ge threshold \\ 0 & otherwise \end{cases}$, respectively. However, the **min** function

needs to be small and the **max** function needs to be large making it difficult to interpret and combine. To address these differences between **min** and **max** FoMs, the FoMs **min** optimal can be inverted such that each the FoMs become **max** optimal, so the function becomes $Optimal = \max(FoMs) + c$. The individual FoMs have different scales, so normalizing is required:

$$\max Normal(x) = \frac{\max(x)}{\max possible\,value} \rightarrow [0..1.0]$$

$$\max(Mission)_i = \sum_{j=0}^{n} w_j FoM_{ij} + c$$

where *w* is the subjective weight of the FoM and $\sum w = 1$ and c represents external information available to the operator.

However, another pattern emerged through the interview – a threshold point in which the operator was unsatisfied with the solution. This threshold point was not a specific value like 10, but rather was dependent on the FoM units. It might be that 15 AAA missiles were "okay", but 20 were not. Likewise, 20 minutes flying over enemy territory was subjectively "okay", but 60 minutes were not. To create a display that can represent different units and different thresholds led to considering the Visual Thinking Sprocket. To address the different units, each FoM was normalized to be into the range of [0..100]. Furthermore, the function was designed to be a two-piece linear function that represented by the linear function [1..threshold] and (threshold..100].

Further bits of fuzziness that cannot be modeled with a computer program are the esoteric influences, such as political, social and economic influences. (For example, do not fly over the site of a "spontaneous" protest march. Avoid state X because a coalition force is active there today. The superior, more experienced officer says not to choose alternative B because "it doesn't feel right".)

After gathering the data, the operator must analyze the resulting parameters to determine if the new missions are acceptable and/or optimal (Figure 4.6). This that implies an informal risk analysis be performed on each alternate mission. Once again, the operator's experience tends to dictate the speed and accuracy of the analysis.

Since the experiment examines an existing complex, dynamic decision support domain, UAV mission stakeholders were interviewed about what data was important to the operator when making mission decisions. These discussions directly led to the eight FoM used in the first two experiments. Furthermore, as would be expected, there was a wide range in the number and subjective priority of FoMs to use, so the sprocket was designed with the flexibility of displaying between six and fifteen FoMs with operator defined weighting.

4.1.2 Research Questions for Decision Support

The objective of this experiment was to examine the viability of applying visual thinking concepts to designing static decision aid user interface displays. Objective data was automatically acquired during the experimental trials, while subjective data was collected through a short questionnaire given to each subject after each trial. Table 4.2 presents the research questions addressed and their associated 1-tailed hypotheses.

Table 4.2 Research questions for Experiment 1

Experiment	Research Question	Related Hypothesis
Accuracy 1	Is there a significant difference in accuracy to obtain good rank order solutions using cognitively sensitive supervisory control of multiple UAVs?	H_0: There was no significant difference in accuracy to obtain good rank order solutions using cognitively sensitve human supervisory control of multiple UAVs. H_1: It was significantly more accurate to obtain good rank order solutions using cognitively sensitve human supervisory control of multiple UAVs.
Accuracy 2	Is there a significant difference in accuracy to determine global status information using cognitively sensitive supervisory control?	H_0: There was no significant difference in accuracy to determine global status information using cognitively sensitive supervisory control. H_1: It was significantly more accurate to determine global status information using cognitively sensitive supervisory control.
Accuracy 3	Is there a significant difference in accuracy to envision new solutions using mental visualization?	H_0: There was no significant difference in accuracy to envision new solutions using mental visualization. H_1: It was significantly more accurate to envision new solutions using mental visualization.
Accuracy 4	Is there a significant difference in accuracy to determine local status along a specific dimension using cognitively sensitive supervisory control?	H_0: There was no significant difference in accuracy to determine local status information using cognitively sensitive supervisory control. H_1: It was significantly more accurate to determine local status information using cognitively sensitive supervisory control.
Response Time 1	Is there a significant difference in time to obtain good rank order solutions using cognitively sensitive supervisory control of multiple UAVs?	H_0: There was no significant difference in time to obtain good rank order solutions using cognitively sensitive human supervisory control of multiple UAVs. H_1: It took significantly less time to obtain good rank order solutions using cognitively sensitve human supervisory control of multiple UAVs.
Response Time 2	Is there a significant difference in time to determine global status information using cognitively sensitive supervisory control?	H_0: There was no significant difference in time to determine global status information using cognitively sensitive supervisory control. H_1: It took significantly less time to determine global status information using cognitively sensitive supervisory control.
Response Time 3	Is there a significant difference in time to envision new solutions using mental visualization?	H_0: There was no significant difference in time to envision new solutions using mental visualization. H_1: It took significantly less time to envision new solutions using mental visualization.

| Response Time 4 | Is there a significant difference in time to determine local status along a specific dimension using cognitively sensitive supervisory control? | H_0: There was no significant difference in time to determine local status information using cognitively sensitive supervisory control. |
| | | H_1: It took significantly less time to determine local status information using cognitively sensitive supervisory control. |

Each probe question used in the Decision Support Experiment was designed to specifically address two of these research questions. For example, Rank Order the three alternate missions from best to worst is designed to examine research question Accuracy 1: **Is there a significant difference in accuracy to obtain good rank order solutions using cognitively sensitive supervisory control of multiple UAVs?** and research question Response Time 1: **Is there a significant difference in time to obtain good rank order solutions using cognitively sensitive supervisory control of multiple UAVs?**

4.2 *Experiment 2: System Monitoring Research Framework*

Following the Experiment 1 success, the same research framework (Figure 4.3) was applied to Experiment 2, namely McNeese, Bautsch *et al*. (1999). This time, the goal (Block 1) was to examine the Visual Thinking Widget as a dynamic system monitoring display within the multi-UAV control domain. This experiment incorporated the Visual Thinking Widget into the MAGES simulation testbed, a medium/high fidelity simulator (Block 2). The subjects were directed to monitor three UAV's gauges simultaneously, determining whether a UAV's "health" was out-of-tolerance. Throughout the experiment, the operator/supervisors were queried about the current and possible future states of the system (Block 3). Following each 20-minute trial, the subjects were surveyed on their opinions of the presented design. So once again, the

subjects were queried during and between experimental trials with SA guided questions (Blocks 5 & 6).

Exp. 2
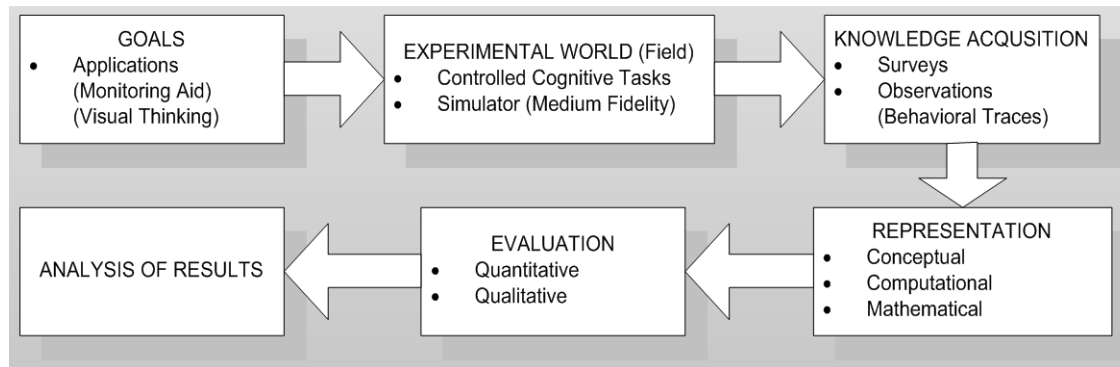
Figure 4.7 Framework for Experiment 2

The experiment follows the monitoring tasks similar to those presented by Narayanan, Ruff *et al*. (2000), but much simpler. The only three options presented to the student subject were 1) do nothing (keep monitoring); 2) identify out-of-tolerance FoM (reboot FoM and note failure); and 3) identify FoMs that may "soon" go out-of-tolerance (watch more diligently and note potential failure). The actions performed by the subjects were pointedly simple, but allowed the subjects to concentrate on the monitoring the system.
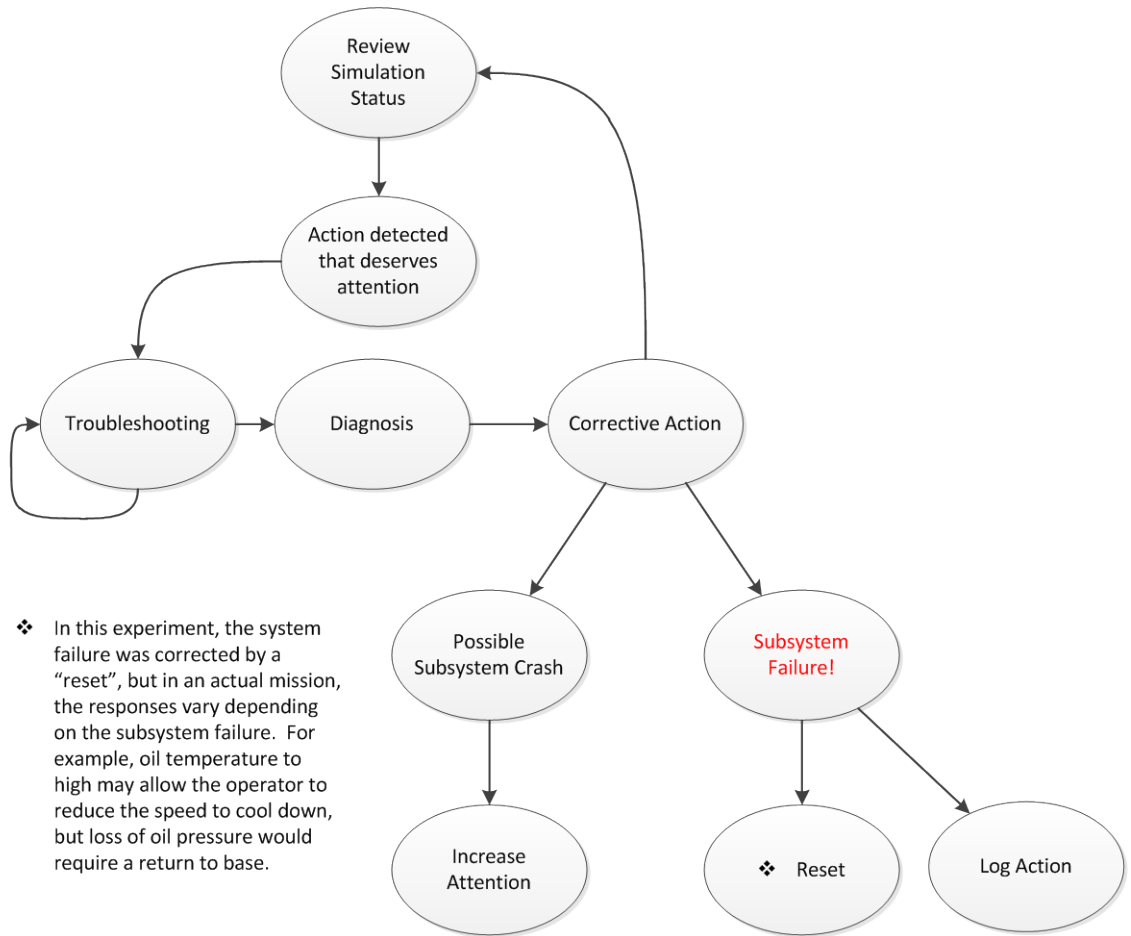
Figure 4.8 OFM of experiment 2 monitoring task

4.2.1 Research Questions for System Monitoring

The objective of this experiment was to examine the viability of applying visual thinking concepts to designing dynamic monitoring user interface displays. Objective data was automatically acquired during the experimental trials, while subjective data was collected through a short questionnaire given to each subject after each trial. The simulation was performed on a medium-high fidelity testbed. Table 4.3 presents the research questions addressed and their associated 1-tailed hypotheses.

Table 4.3 Research questions for Experiment 2

Experiment	Research Question	Related Hypothesis
Distracter: Root Mean Square Error [RMSE]	Can the operator perform the primary task (keeping the target centered) better using cognitively sensitive human supervisory control of multiple UAVs? [RMSE]?	H_0: There was no significant difference in RMSE in keeping the target centered H_1: There was significantly less RMSE in keeping the target centered.
Monitor Task: Accuracy	Is there a significant difference in accuracy when monitoring out of tolerance values?	H_0: There was no significant difference in accuracy when monitoring out-of-tolerance values. H_1: There was significantly more accurate when monitoring out-of-tolerance values.
Monitor Task: False Positive	Is there a significant difference in false positives when monitoring out-of-tolerance values?	H_0: There was no significant difference in false positives when monitoring out-of-tolerance values. H_1: There was significantly more false positives when monitoring out-of-tolerance values.
Response Time	Is there a significant difference in time to respond to out-of-tolerance monitored values using cognitively sensitive supervisory control?	H_0: There was no significant difference in time to accurately respond to out-of-tolerance monitored values using cognitively sensitive human supervisory control of multiple UAVs. H_1: It took significantly less time to accurately respond to out-of-tolerance monitored values using cognitively sensitive human supervisory control of multiple UAVs.

4.3 *Experiment 3 Research Framework Details*

Still following the flow of the McNeese framework, the final experiment initially was to instrument the existing Predator simulation software with the Visual Thinking Widget library, and then have active duty and reserve UAV operators use the system. However, due to internal military and external corporate requirements[11], the validation

---

[11] The copyright to the Predator Simulation software is held by a subcontractor of the simulation system. While the primary contractor had said that the instrumentation could be performed on the simulator, it was contingent on permission being granted by the primary copyright holder. That decision on granting

was achieved using the existing MAGE software along with the active duty and reserve

UAV operators.

Exp. 3
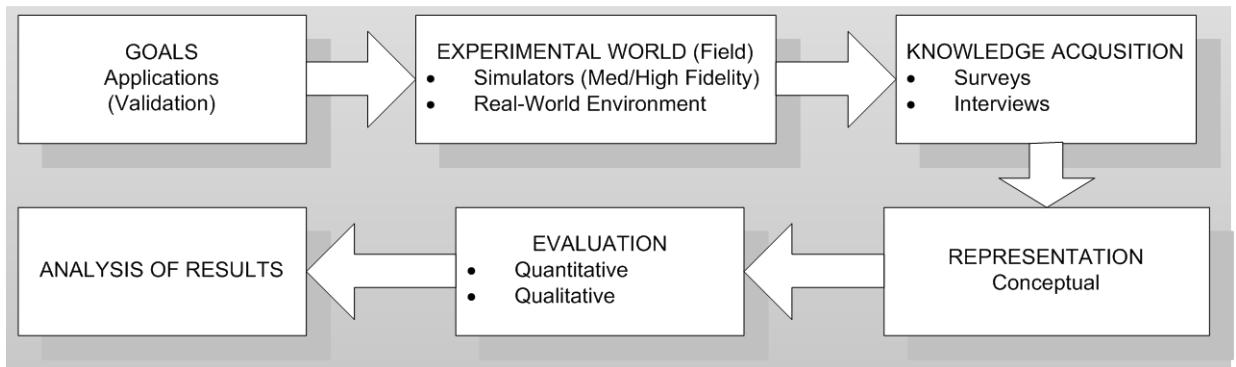
Figure 4.9 Framework for Experiment 3

During the experiment, the UAV operators were surveyed and interviewed about

the various MAGE training enhancements. These included the Visual Thinking displays

(decision aid and monitoring system) along with other enhancements[12]. These interviews

were held in Victorville, CA with the Predator training site and at Wright Patterson AFB

Human Effectiveness Laboratory.

---

the permission was postponed until one week prior to the experiment, at which time the primary
copyright holder decided to **not** grant permission. At that time, it was decided to have the UAV
operators use the MAGE software to grant permission. So it was decided to have the UAV operators
observe the MAGE software to validate the concept.

[12] The other enhancements included a mission timeline decision aid, network centric information displays,
real time weather low fidelity modeling,

4.3.1 Research Questions for third Experiment

Table 4.4 Research questions for Experiment 2

	Research Question	**Related Hypothesis**
Workload preference	Does the SME prefer the VTS to the current or modified VIT system?	H_0: There was no significant difference in subjective workload preference among the three displays
		H_1: There was significant difference in subjective workload preference among the three displays

5 Experiment Design

5.1 *Decision Aid Support Experiment*

The experiment presented three alternate static displays in a controlled laboratory environment. The displays were to be used as decision aids for UAV mission planning. This experiment's goal with respect to this research was intended to answer whether the visual thinking widget was an effective display alternative as a decision aid.

5.1.1 Training

Training consisted of 30 minutes of instructions via an oral PowerPoint® presentation, and visually (projected on a wall mounted screen and on each subject's computer desktop). Furthermore, each subject was given a paper copy to refer to "as needed" during the experiment. Each of the three display formats were examined in detail, with each prompting question examined along with verbal descriptions of what the interpretations should be.

5.1.2 Design

The experiment was a two factor repeated-measures design with full-model partitioning. Microsoft Excel® was used to generate the text table and bar chart images that were compared with the sprocket, while software developed for the MAGE Decision Support System (DSS) generated the sprocket images. Each image was generated from the same data, i.e., a bar chart, text table and sprocket image were generated from dataset 1, 2, 3, etc. Each subject was shown a series of generated images of the Figures of Merit

(FoM) for three alternative paths' and a question. The subject responded to the first question (rank order the three routes from best to worst), and then the second question was displayed, and so on – the presented image did not change and remained visible during the questioning (SPAM – Situation Present Assessment Method). The four questions, possible answers, levels of situation awareness, and question abbreviation are listed in Table 5.1. The question abbreviation is used in the Results section as an easy mnemonic for each of the question types.

Table 5.1 Experiment 1 (Static Displays) questions

	Situation Awareness Level	Question	Possible Answers	Abbreviation
1	Level 2 SA	Rank order the routes [Best to worst]	1-2-3 1-3-2 2-1-3 2-3-1 3-1-2 3-2-1	Rank_Order
2	Level 1 SA	Do any of the routes meet the all minimum criteria?	Yes No	Minimum_Criteria
3	Level 3 SA	Which route is Best if Dimension X is dropped? (Where X was chosen from among the 4 top weighted available dimensions)	1 2 3	Drop_Dimension
4	Level 1 SA Level 2 SA	Which route has Best Dimension Y? (Where Y was chosen from all available dimensions)	1 2 3	Best_Dimension

After the fourth question, a new image was displayed and the same four questions were asked. For each display type, 12 images were generated with the same datasets. In other words, 12 datasets were used to generate 12 bar chart, table and sprocket images for a total of 36 images. Referring to Table 5.2, the two factors are **display type** (text table, bar chart, and sprocket) and **image dataset** (I_1, I_2, …, I_{12}). The order of presenting the displays was fully-balanced and randomly assigned to each subject to avoid learning

effects. The order of presenting the images was also randomly assigned per block; however, there were only three lists of random images. Each subject saw 36 images and answered four questions for each image for a total of 144 questions. For each question, the response time and answer were recorded, thus giving two objective measures for each question.

Table 5.2 Outline of the design with two repeated-measures factors

		Images (I=12)											
		I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}	I_{12}
Display (D=3)	Table D_1	I_1D_1	I_2D_1	I_3D_1	I_4D_1	I_5D_1	I_6D_1	I_7D_1	I_8D_1	I_9D_1	$I_{10}D_1$	$I_{11}D_1$	$I_{12}D_1$
	Bar Chart D_2	I_1D_2	I_2D_2	I_3D_2	I_4D_2	I_5D_2	I_6D_2	I_7D_2	I_8D_2	I_9D_2	$I_{10}D_2$	$I_{11}D_2$	$I_{12}D_2$
	Sprocket D_3	I_1D_3	I_2D_3	I_3D_3	I_4D_3	I_5D_3	I_6D_3	I_7D_3	I_8D_3	I_9D_3	$I_{10}D_3$	$I_{11}D_3$	$I_{12}D_3$

Each mission route has seven weighted Figures of Merit (FoM) in which the subject was to judge the routes (Table 5.3). Each FoM was presented to the subject in a predetermined order primarily based on its weighting value. Although the FoM measures would likely have some interdependencies (*i.e.*, SAM shots and SAM tracking probability of detection have an obvious relationship), their test values were randomly generated and independently determined.

Table 5.3 Mission Figures of Merit descriptions and order of presentation

Figure of Merit	Weight	Description
Probability of Survival	40	Self-explanatory
Surface-to-Air Missile (SAM) shots	15	Number of SAM missile launch opportunities by the enemy
Aircraft Intercept launch	15	Number of aircraft intercept opportunities (determined by tracking) that can intercept the UAV
Anti-Aircraft-Artillery (AAA) exposure	15	Number of minutes the UAV is exposed to enemy AAA
Net number of tracks exposed	9	For each track (leg) of the route, how many allowed radar tracking
SAM tracking probability of detection	3	Given a SAM battery, the estimated probability of detection
Early Warning (EW) radar probability of detection	3	The possibility of an early warning system detecting the UAVs entering the enemy airspace

117

Each block of displays had twelve images of three alternative mission routes. Figure 5.1 illustrates the same three routes displayed as a table, bar chart and a visual thinking sprocket image. In the Table Display format (Figure 5.1.A), the first column lists the weight given for each FoM, the second column states the FoM's name – so the Probability of Survival is the most important dimension reflected in the 40%, while SAM tracking probability of detection, Early Warning (EW) radar probability of detection share the bottom at a 3% weighting value. The third, fourth and fifth columns represent the three alternative routes. The sixth and seventh columns are used in conjunction to assign pass/fail for the specific FoM. The sixth column states the minimum/maximum threshold for that respective FoM, any value failing to meet this threshold is colored red. The last column refers to the directionality of the FoM axis: the plus sign (+) means bigger is better, while the minus sign (-) means smaller is better. (Only the Probability of Survival has a positive direction, *i.e.*, the value must be larger than the criterion to pass. All of the other FoM require smaller values to pass the criterion.)

Looking at the first route (3rd column) in Figure 5.1.A, one can see that for the Probability of Survival (2nd row) – "bigger is better" (7th column) and fails to exceed the minimum criterion at 0.11 (6th column), and is thus colored red (3rd column).

The Bar Chart Display (Figure 5.1.B) and the Visual Thinking Sprocket Display (Figure 5.1.C) formats supply the same information as the Table Display. (See Appendix A for more complete description to the experiment.)

Weighting	Figure of Merit	Route 1	Route 2	Route 3	Criterion	Better
40	probabilityOfSurvival	0.11	0.84	0.46	0.2	+
15	samShots	1	4	11	10	-
15	aiLaunch	4	9	9	8	-
15	aaaExposure	927	670	805	300	-
9	netNumTrack	15	10	5	15	-
3	samTrackingRadarPd	0.6	0.33	0.76	0.6	-
3	ewRadarPd	0.97	0.16	0.58	0.8	-

A) Table display
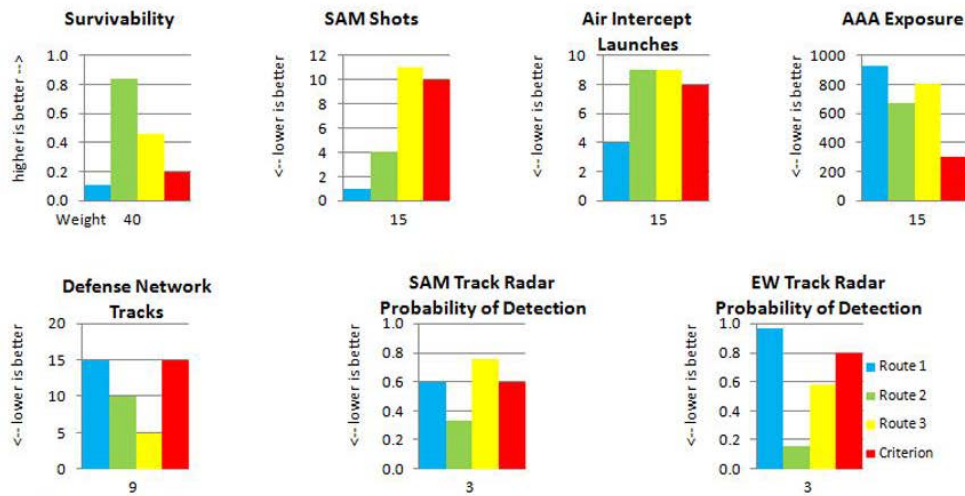
B) Bar chart display
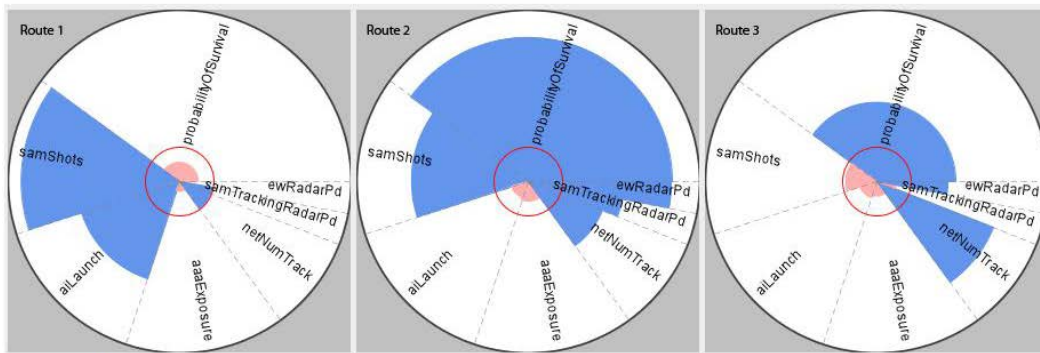
C) Visual thinking sprocket display

Figure 5.1 Three different display modes of the same mission alternatives

The same twelve data profiles were used in each image, but the profiles were presented in six different orders to reduce carry over effects. This was so the subjects

were not able to detect similarities across blocks. The blocks were counter balanced to further reduce any carry over effects.

5.1.3 Participants

The research design and procedures were submitted to the Wright State University Institutional Review Board, who approved conduct of the study as a low-risk protocol. The first experiment had 24 volunteer subjects, mostly engineering graduate students, recruited from Wright State University. All subjects saw the same images. The order of presentation was varied by display type (3) and mission/route order (12) using a Latin Square design. Each subject was given a Subject Consent Form prior to starting the experiment. This was followed by a thirty minute PowerPoint® presentation of background material and instructions for the experiment. Training, presentation of the three conditions, and debrief of the subjects took a total of 2.5 hours.

5.1.4 Evaluating Rank Order

This research explores a novel approach to grading the rank-order selection of the by the subjects (question 1 above). A correct/not correct grading seems not a strong enough statistical measure of wrongness, nor is it a fine enough measure of the permutations of wrong. Referring to Table 5.4, I propose using a weighting function that places a premium on the number of orderings that are correct. For instance, if the correct ordering of the routes was 1-2-3, then one can break this ordering into three pairs: 1 before 2, 2 before 3 and 1 before 3; henceforth written as 1→2, 2→3 and 1→3, respectively. One then considers the six permutations of these the original three numbers

and counts the number of correct paired orderings: the first column is 1-2-3, and has three correct pairings ($1\rightarrow2$, $2\rightarrow3$ and $1\rightarrow3$), giving us a correct ordering value of 3; the second column has two correct pairings ($1\rightarrow2$ and $1\rightarrow3$), giving a value of 2; the third column has two correct pairings ($1\rightarrow3$ and $2\rightarrow3$), giving a value of 2; etc. In short hand notation, if 1-2-3 were correct, then one could write Selection_{123} ($1\rightarrow2$, $2\rightarrow3$, $1\rightarrow3$) = 3; Selection_{132} ($1\rightarrow2$, $1\rightarrow3$) = 2, Selection_{213} ($2\rightarrow3$, $1\rightarrow3$) = 2, Selection_{231} ($2\rightarrow3$) = 1, Selection_{312} ($1\rightarrow2$) = 1, and Selection_{321} (ø) = 0. A Table 5.4 lookup table is created for each of the six possible permutations of ordering truths. The benefit of this function is that the answers are objectively partitioned into four groups of relative wrongness [3, 2, 1, 0], giving us a way to measure the "wrongness" of an answer.

Table 5.4 Evaluating rank order

Truth	Order Selected by Subject					
1	1	1	2	2	3	3
2	2	3	1	3	1	2
3	3	2	3	1	2	1
Ordinals correct	3	2	2	1	1	0

An alternate weighting function in which the correct placement of the first two items in the answer is rewarded with bonus points: first (+2) or second (+1) correct. For instance, in the first column, since 1 is in the first place setting this answer gets a bonus of +2, and since the 2 is in the second place setting this answer gets a bonus +1; for a total bonus of +3 and grand total of 6. The second column has the 1 in the correct place but not the 2, so it gets a bonus of +2 points. Columns 3, 4 and 5 have no bonus points, but column 6 gains a bonus point. This weighting function was not used since the subjects were not told to concentrate on getting the best first (and because it adds some undeserved ambiguity to the "worst" three choices). [Again, assuming that 1-2-3 was the

correct order, the final scores with any bonuses added would be: Selection$_{123}$ (1→2, 2→3, 1→3,+2,+1) = 6; Selection$_{132}$ (1→2, 1→3,+2,0) = 4, Selection$_{213}$ (2→3, 1→3,0,0) = 2, Selection$_{231}$ (2→3,0,0) = 1, Selection$_{312}$ (1→2,0,0) = 1, and Selection$_{321}$ (ø,0,1) = 1]. Again we have 4 levels of discrimination.

5.2 *System Monitoring Experiment*

The experiment presented three alternate animated displays in a controlled laboratory environment. The displays were to be used for system monitoring of four simulated UAVs. Since the first experiment was a success, this experiment's goal with respect to this research was intended to answer whether the visual thinking widget was an effective display alternative as a system monitoring aid. To this end, the visual thinking widget could then be considered a flexible alternative to other displays.

5.2.1 Training

Training consisted of 30 minutes of instructions via a PowerPoint® presentation, the presentation was made with a projector and a copy of the PowerPoint® presentation was on each subject's computer desktop so they could follow along. Each subject received a color copy of the PowerPoint® slides that they could keep for reference during the experiment. Each display format was examined in detail, with each prompting question examined along with verbal descriptions of what the interpretations should be.

Following the PowerPoint® slides, the subjects were given a 5 minute limited response training simulation. The screen layout was the same as the actual experiment display, but only 2 of the feedbacks for the sensors were active, i.e., these two sensors

122

initially started near 50 (neutral), went out of tolerance, and clicking on those sensors reset the sensor to a neutral value. The only sensor display used in the training simulation was the analog sensor. The training simulator was primarily intended to give the subjects experience in keeping the target in the crosshair sights. The sensor gauges that were live were explicitly specified to the subjects, and what "going out of tolerance" meant was explained.

5.2.2 Design

The experiment was a two repeated-measures factors design with full-model partitioning. Java® standard software libraries were used to generate the text and gauge table animations, while software developed in Java® for the MAGE decision support system generated the sprocket animations. The experiment was divided into two tasks: the primary task was to maintain target acquisition within the camera's crosshair while the UAV drifts (because of simulated turbulence and wind) and the secondary task was monitoring three UAV systems. (See Section 5.2.4 for a description of the experiment flow.)

Microsoft Excel® was used to generate the data used in the animated text table, analog gauges and sprocket displays, guaranteeing consistent repeatable experiments. Each animation was generated from the same data, i.e., the gauge, text table and sprocket images were generated from dataset 1, 2, 3, etc. Each FoM had over 1,500 data points to ensure coverage of at least a 25 minute simulation time. Since this was interspersed with questions that froze the simulation clock, the data was sufficient for 30 minute trials.

Table 5.5 Experiment 2 (Dynamic Displays) questions

Situation Awareness Level	Question	Answer Type
Level 1 SA	For UAV_Y, what is the value of the $Gauge_X$?	Slider
	Did UAV_Y stay in tolerance?	Boolean
	Did UAV_Y have a Gauge go out of tolerance?	Boolean
Level 2 SA	Is the $Gauge_X$ in UAV_Y stable? (+/-10 from center)	Boolean
	How many UAVs are currently healthy?	Count
	Which UAV is MOST healthy? (Closest to center)	Selection
	Is the $Gauge_X$ in UAV_Y at the desired level?	Boolean
Level 3 SA	Which UAV is most likely to fail? (Maintenance Report)	Selection
	Which UAV will go out of tolerance next?	Selection
	Can the $Gauge_X$ for UAV_Y exceed tolerance in the next 5 seconds?	Boolean

5.2.3 Participants

The second experiment had 24 subjects, recruited from Wright State University graduate students. All subjects saw the same animations, however, the order of presenting the sensor displays varied for each animation and the data behind the animations was randomly assigned to the display type.

5.2.4 Experiment flow

As mentioned earlier, the two tasks were to maintain target acquisition within the camera's simulated crosshair while the UAV drifts (because of simulated turbulence and wind) and monitoring four UAV systems (See Figure 5.2).
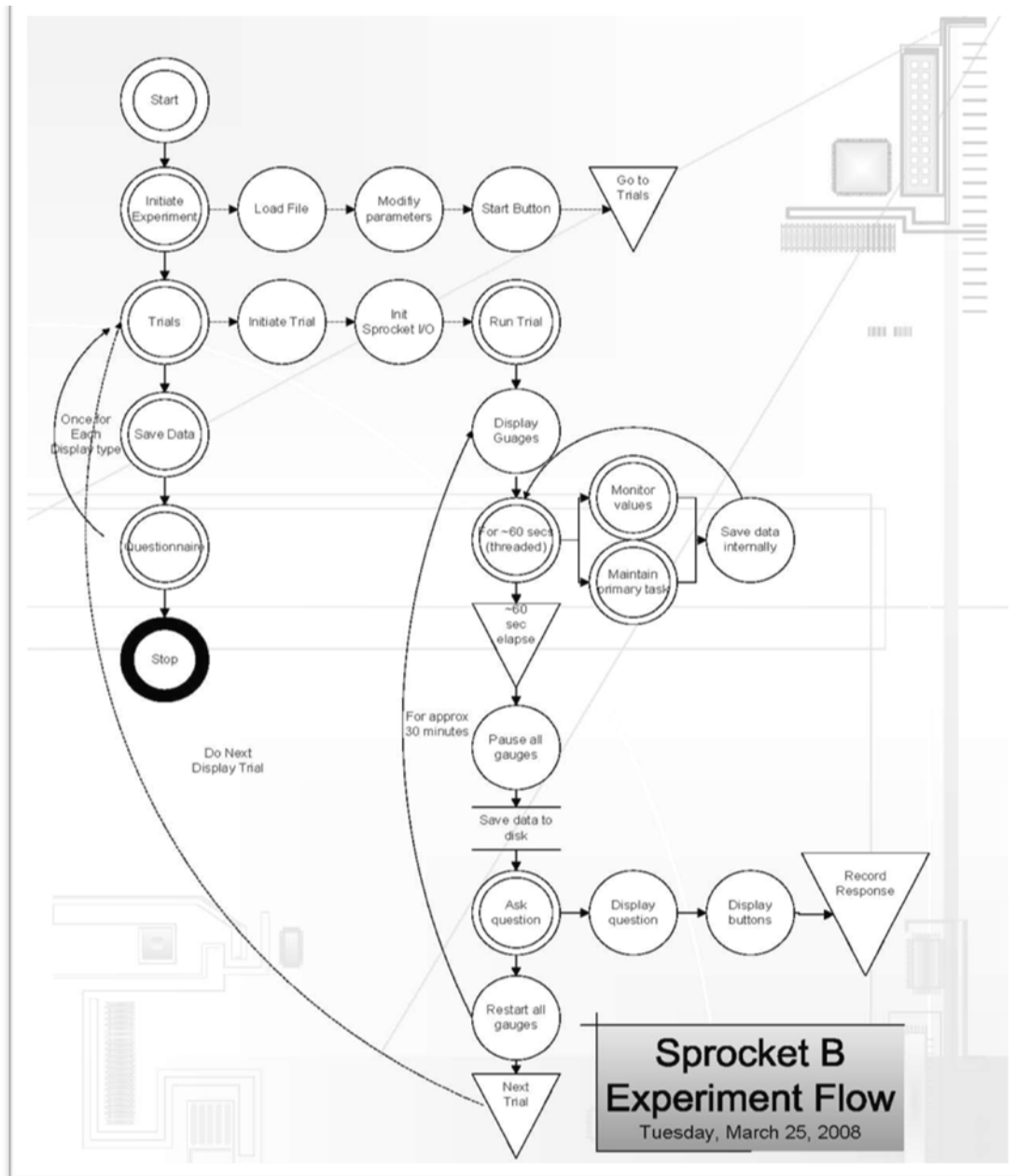
Figure 5.2 System monitoring experiment flow

The right hand column in Figure 5.2 is an overview of each trial – initialize the data, run the trial, present a questionnaire, save any additional data, repeat. During the "run the trial" portion, the subject was expected to perform these two tasks, while intermittently answering questions. The primary task was to control the UAV's camera to be trained on a targeted building. The targeted building is represented by a red box

surrounding a building on one of three aerial photographs of Bagdad, Fallujah and Hadid. Each subject was also shown animations of simulated monitored systems and was instructed to respond by clicking on the gauge if it goes out of tolerance. The initial conditions of the monitored systems place all the gauges in-tolerance and they "drift" according to the scripted datasets.

During the test, the subject had to answer probe questions (*e.g*., How many times were you aware of out of tolerance conditions? Is UAV_1 completely in tolerance at this moment? *etc*.) The systems monitoring experiment examines an animated Visual Thinking Sprocket with two criteria (thresholds) per dimension for monitoring the health of multiple UAVs. Each of these Visual Thinking widgets were designed to be flexible in number of dimensions (5-15) and different weighting values for each dimension. However, for this experiment, six dimensions were chosen.

Each subject was shown a series of generated animations for three UAVs, every 60-120 seconds a question was presented to the user. The variation in presenting the times was to avoid the situation where the subject anticipates/times the question interval and changes their intensity of the concentration. When the subject responded to a question, while the animation was halted, the presented image did not change and remained visible during the questioning (SPAM – Situation Present Assessment Method). The three levels of situation awareness, ten associated questions, and their corresponding input mechanism are listed in Table 5.5. During each pause, a set of three questions were asked, one from each SA level. Which question was asked was randomly chosen and dynamically populated with appropriate FoM and UAV data. For example, Level 1 SA

question "For UAV_Y, what is the value of the $Gauge_X$?" could be asked as "For UAV_1, what is the value of the Air Temperature Gauge?" or "For UAV_2, what is the value of the Engine RPM?"

5.3 *Full system experiment*

The experiment was a usability study using subject matter experts (SMEs) from Victorville, CA and Wright Patterson AFB, OH with concurrent protocols and examined a Visual Thinking decision support aid in the existing full MAGE simulation testbed and in particular it's Decision Support System (DSS).

5.3.1 Training

Compared to the existing Predator control mechanisms, operator responses are different in the MAGE system. Currently, the Predator is manually flown and largely manually controlled. MAGE on the other had is semi-autonomous in which operators exercise supervisory control. These SMEs are familiar with many of the operational issues of UAV operations and are best qualified to appreciate how those issues interact with UAV operations. Their instructions were to apply their knowledge to the supervisory control scheme, although they were free to comment and criticize the elements and assumptions of the MAGE system.

5.3.2 Design

Usability testing lasted approximately 2.5 hours. During the first hour, the participant completed an UAV experience questionnaire that captured their qualifications as a Subject Matter Expert (SME). This hour included approximately 45 minutes of

familiarity training with the MAGE system. A short demonstration scenario was run showing each of the typical events the SME would experience in the longer test scenario. They were familiarized with the workload and other assessment methods employed during the usability testing.

The test scenario was specifically designed to exercise all aspects of MAGE, with particular emphasis on the DSS system. The scenario requires monitoring and control over two UAVs flying mixed Intelligence, Surveillance and Reconnaissance (ISR) and target strike missions over Iraq. The missions were choreographed to contain event conflicts between the UAV missions to see how well the DSS mitigates the temporal conflicting mission demands. The operator received mission updates from higher authority which added or decreased tasking during the different mission segments. Furthermore, the displays used during a specific leg of the mission were automatically changed to visit all three display types.

Besides the automated action data logging, several other measures were used during the test scenario. First, audio recording were made of each session and the subjects were instructed during training to "Think Out Loud", verbalizing their impressions of the system during the scenario. These audio recordings were screened for important observations, and carefully scrutinized during key events to extract the most information about the DSS effectiveness.

Second, the National Aeronautics and Space Administration Task Load Index (NASA TLX) was used to capture workload assessment between missions. The SME's performed a sample NASA TLX during their study preparation. It should be noted that

the NASA TLX used in this study in known as Raw NASA TLX (Hart 2006). It eliminates the weighting process all together or weighting the subscales and then analyzing them individually. The subjects do not have to do a pairwise ranking of the six scales (Mental, Physical, Temporal, Frustration, Effort and Performance). The subjects were instructed to use the GCS Variable Information Tables (VIT) as a baseline definition of the scale. That is, assume the mental/physical/temporal/effort/frustration/performance scales for the VIT is 50%. If that were the case, how would the other displays be rated compared to that baseline.

Figure 5.3 NASA TLX program interface (Sharek 2010)

The last section of the study session was used to complete a close-out questionnaire about MAGE's function and to collect open-ended verbal comments from the SMEs. These data was comprised of the SME subjective evaluation of the DSS and MAGE.

5.3.3 Participants

Five current Predator Pilots (P) or Sensor Operators (SO) served as SME participants for the DSS Usability Study. These P/SO are civilians employed by USI, Inc. in Victorville, CA. They are former military P/SOs with operational experience, who now flies the Predator for their employer in support of operational test and training at Edwards AFB, CA. Furthermore, three additional P/SO SME participants were used from WPAFB. The P/SOs were active duty or civilian contractors that were at WPAFB for a three day system review of the Predator system. A background survey captured each participant's exact experience with UAV operations.

5.4 *Summary*

Each experiment presented in this research was carefully designed to specifically address whether the visual thinking paradigm could be incorporated into a highly complex, dynamic system. The first showed that the concept was sound, while the second experiment indicated that the concept was generalizable, and finally, the third experiment was to try to understand the barriers and hurdles of incorporating the displays into an existing environment.

The overviews of the three experiments' variables identified are presented in Table 5.6.

Table 5.6 The variables involved in the experiments

	Experiment 1	Experiment 2	Experiment 3
Quantitative	Yes	Yes	No
Independent Variables	Display Type	Display Type	Missions run w/ and w/o baseline net-centric supplemental display
	Questions	Questions	
Dependent Variables	Rank Order	Distracter: RMS Error centering target	Observations
	Response Time$_1$	Monitor Task: Accuracy	Think Out Loud Protocol
	Minimum criteria	Monitor Task: Response Time	NASA TLX scores
	Response Time$_2$		
	Drop FoM **X**		
	Response Time$_3$		
	Best Dimension **Y**		
	Response Time$_4$		
Questions	Questionnaire	Probe questions on current or future state	Exit Questionnaires

6 Results

6.1 *Decision Aid Support Experiment*

The data was comprised of answers to questions related to decision support choices found in the three static displays presented to the operators. The decision support questions were based on mission critical planning information called Figures of Merit (FoM) – see Table 4.2. The data collected for this experiment used questions to ascertain the ease of use as a decision support tool for each display. Recall that the text table represents the current UAV display technology (circa 2007) used in a Predator Ground Control Station (GCS) Variable Information Tables (VIT). The bar chart was chosen as a first naive attempt at a "graphical user interface". It is understood that these are discrete interfaces, not integral interfaces, which is because there exist no integral interfaces for multiple UAVs.

The displays were static images, meant to elicit quickly identified solution patterns for the operator. The questions were designed to explore specific decisions considered typical within a multi-UAV mission.

6.1.1 Analysis of Data

The data for the operators' answers to the static displays were analyzed using Statistical Analysis Tool® (SAS v 9.2) for windows. There were 24 subjects that had three display types of twelve images with four questions each (24 subjects x 3 blocks x 12 images x 4 questions = 864 Total Observations). The operators' responses were

grouped and analyzed by display type (text table, bar chart and sprocket), for correctness of answer per question type (Rank_Order, Minimum_Criteria, Drop_Dimension, and Best_Dimension), response time per question type and response over time (learning curve).

6.1.1.1 Correctness

Looking at the "correctness of answer" per question data first, the generated experiment data were examined prior to presentation to the subjects to determine the "correct answer" to each question. The operator's "correctness" response was then defined as whether the operator responded with the pre-calculated correct answer. A correct response was assigned a value of 1 and a wrong answer was assigned a value of 0. The sum of the correct answers was then used as a measure to determine the cognitive ease of use for each display type – each question had a maximum score of 12 points for correctness.

For example, the Minimum_Criteria question asks whether all of the mission alternatives presented to the operator meet the minimum criteria on all dimensions: (1) the possible answers are Yes or No, and (2) the correct answer is Yes (the three mission statements are all valid).[13] Then for each operator that answered Yes, a counter would be incremented by 1. The maximum value the counter could reach is 12, so if the

---

[13] The Text Table and the Sprocket color coded a failure of each dimension, i.e., if the "Probability of Success" dimension failed to reach the minimum value, then the text (pie piece) would be colored red. For the Bar Chart, no color coding was attempted.

final counter value was 12 (out of 12), then 100% of the subjects responded with the correct answer. If ½ of the subjects responded correctly, then the final value would be 6 (out of 12).

The results of the one-way repeated measures ANOVA are shown in Table 6.1.

Table 6.1 Tests for Correctness of Within-Subjects Question Type with Subject*Display as an Error Term

	Source	DF	ANOVA SS	Mean Square	F Value	Pr > F	Sig
Rank_Order (Boolean)	Display	2	66.08	33.042	11.70	<.0001	*
Error Term	Subject*Display	46	129.92	33.042			
Minimum_Criteria	Display	2	24.69	12.347	7.82	0.0012	*
Error Term	Subject*Display	46	72.64	1.579			
Drop_Dimension	Display	2	50.78	25.389	14.03	<.0001	*
Error Term	Subject*Display	46	83.22	1.809			
Best_Dimension	Display	2	10.75	5.375	2.03	0.1432	
Error Term	Subject*Display	46	121.92	2.650			
Rank_Order (Levels of Correctness)	Display	2	182.25	91.125	10.40	0.0002	*
Error Term	Subject*Display	46	403.08	8.763			

Interpreting Table 6.1 is straight forward: For line 1, there exists a significant main effect of Displays [$F(2, 46)=11.70$, $p<0.0001$] when judging the Rank_Order of the three missions using simple Boolean correct/incorrect scoring. *Post hoc* analysis (Table 6.2) using the Tukey method for multiple comparisons with an alpha level set to $\alpha=0.05$ revealed that operator responses with the Sprocket were significantly more accurate than either the Bar Chart or Text Table.

Table 6.2 Rank_Order BOOLEAN Tukey Correctness grouping (Means with the same letter are not significantly different)

```
Tukey Grouping   Mean    N    Block
        A        7.17    24   Sprocket
        B        5.38    24   Bar Chart
        B        4.96    24   Text Table
```

For line 2, Table 6.1 shows there exists a significant main effect of Displays [$F(2, 46)= 7.82$, $p=0.0012$] when judging the Minimum_Criteria. *Post hoc* analysis (Table

6.3) using the Tukey method for multiple comparisons with an alpha level set to α=0.05 revealed that operator responses with the Sprocket and Text Table were significantly more accurate than the Bar Chart. This is most likely attributable to the color coding in the Sprocket and Text Table displays: operators were able to immediately determine a bad dimension without having to search the display.

Table 6.3 Minimum_Criteria Tukey Correctness grouping (Means with the same letter are not significantly different)

Tukey Grouping	Mean	N	Display
A	11.29	24	Sprocket
A	10.96	24	Text Table
B	9.92	24	Bar Chart

For line 3, Table 6.1 there exists a significant main effect of Displays [$F(2, 46)=$ 14.03, $p<0.0001$] when judging the Drop_Dimension question. *Post hoc* analysis (Table 6.4) using the Tukey method for multiple comparisons with an alpha level set to α=0.05 revealed that operator responses with the Sprocket were significantly more accurate than either the Bar Chart or Text Table.

Table 6.4 Dimension_Drop Tukey Correctness grouping (Means with the same letter are not significantly different)

Tukey Grouping	Mean	N	Display
A	9.17	24	Sprocket
B	7.75	24	Bar Chart
B	7.17	24	Text Table

For line 4, Table 6.1 shows NO significant main effect of Displays [$F(2, 46)=$ 10.75, $p=0.1432$] when judging the Best_Dimension.

For line 5, Table 6.1 shows there exists a significant main effect of Displays [$F(2, 46)=$ 10.40, $p=0.0002$] when judging Rank_Order of the three missions using Levels for correct/incorrect scoring. *Post hoc* analysis (Table 6.5) using the Tukey method for multiple comparisons with an alpha level set to α=0.05 revealed that operator responses

with the Sprocket were significantly more accurate than either the Bar Chart or Text Table.

Table 6.5 Rank_Order LEVELS Tukey Correctness grouping (Means with the same letter are not significantly different)

Tukey Grouping	Mean	N	Block
A	29.88	24	Sprocket
B	26.50	24	Bar Chart
B	26.50	24	Text Table

Unfortunately, comparing the results of Line 1 (Rank_Order – BOOLEAN) with Line 5 (Rank_Order – LEVELS) shows no gain in the strength of the statistical analysis using the more sophisticated measure for Rank_Order. This does not mean that the "Levels of Correctness" is not a valid tool; just that in this circumstance it gives us nothing more.

In summary, the first three question types (Rank_Order, Minimum_Criteria, and Drop_Dimension) showed significant differences in the Correctness for the displays, with the Sprocket display being either the best or tied for best display. In the last question (Best_Dimension), there was no statistical significance in accuracy among any of the displays.

6.1.1.2   Response Time

A two repeated measure ANOVA was used to test the main effects of display (Text Table, Bar Chart, and Visual Sprocket) and Trial (12 images) on the operators' response time for each question type (Rank_Order, Minimum_Criteria, Drop_Dimension, Best_Dimension). The next four subsections analyze these SAS ANOVA results.

Each analysis has two tables. The first of the pair has the calculated ANOVA Sum of Squares (SS) and Mean Square of the sources of error generated by SAS: The

sources of variance are the same for each question type. The second table has the main

effects of Display and Trial and the interaction of Display*Trial analyzed with the

appropriate corrected error.

6.1.1.2.1 Response Time for Rank_Order

The first question examines the operator's ability to mentally rank the three

mission alternatives from best to worst based on the FoM information presented in each

display.   The operators' response time needed to Rank_Order the alternate mission

images is analyzed.

Table 6.6 Rank_Order Response Time Tests of Within-Subjects Effects

	Source	DF	ANOVA SS	Mean Square	F Value	Pr > F	Sig
Effect	Display	2	98420818853	49210409427	30.91	<.0001	*
Error	Subject*Display	46	73242779526	1592234338			
Effect	Trial	11	104379784322	9489071302	17.44	<.0001	*
Error	Subject*Trial	253	137651423089	544076771			
Effect	Display*Trial	22	52837127566	2401687617	7.99	<.0001	*
Error	Subject*Display*Trial	506	152093131835	300579312			

The main effect of Display was found to be significant: [F(2, 46)= 30.91,

p<0.0001]. This means that the Display type used can significantly affect the speed with

which the operator interprets the ranking of the mission alternatives. The main effect of

Trial   (image   presentation   order)   was   found   to   be   significant:

[F(11,253)=17.44, p<0.0001].   Images presented at the beginning of the trial took

significantly longer to interpret than those presented at the end of the trial, irrespective of

display type. This is also known as a "learning curve".

Examining the *post-hoc* Rank_Order Tukey Response Time grouping by Display

Type (Table 6.7) results for multiple comparisons with an alpha level set to α=0.05

revealed that operator responses were significantly faster for the Sprocket Display (smaller is better).

Table 6.7 Rank_Order Tukey Response Time grouping by Display Type (Means with the same letter are not significantly different)

Tukey Grouping	Mean (Sec)	N	Display
A	44.049	288	Bar Chart
B	23.616	288	Text Table
C	19.708	288	Sprocket

Furthermore, examining the *post-hoc* Rank_Order Tukey Response Time grouping by Image Presentation Order (Table 6.8) results for multiple comparisons with an alpha level set to $\alpha=0.05$ revealed that there was a learning effect, i.e., the numbers started large and quickly leveled out (smaller is better). Since the order of the image presentation was random, the learning effect was similar for each display type (generally declining response time). However, response times for the early images showed different levels of cognitive workload for each display type. [For one subject, the first image was presented and then questions were asked about interpretation. These questions had been answered in the training, but the subject's lack of familiarity with the display resulted in an unusually high response time for the response for the very first image presented – over 330 seconds (5.5 minutes). Replacing that data point with the mean of the subject pool does not significantly change the results of the statistical findings: Tukey Grouping for the Display Response Time changed from three to two groups with Text Tables and Sprocket not being significantly different, and the Response Time Ordering for the images goes from 1, 2, 3 to 1, 3, 2 – the groupings remain the same!]

Table 6.8 Rank_Order Tukey Response Time grouping by Image Presentation Order (Means with the same letter are not significantly different)

Tukey Grouping	Mean	N	Trial
A	60084	72	1
B	38708	72	2
BC	35656	72	3
BCD	31358	72	5
CDE	26211	72	7
DE	25612	72	4
DE	24559	72	9
DE	23459	72	6
DE	22908	72	8
DE	22380	72	11
E	19290	72	12
E	19269	72	10

6.1.1.2.2 Response Time for Minimum_Criteria

Next, the main effects on response time for the Minimum_Criteria were examined (Table 6.9). As expected, the type of display had a strong statistical effect on the response time [$F(2,46)=29.86$, $p<0.0001$]. Also as expected, the presentation of images over time improved statistically (had a learning curve) [$F(11,253)=13.27$, $p<0.0001$]. Finally, when considering the Display_Type in conjunction with the presentation of the images showed a statistical difference [$F(22,506)=2.80$, $p<0.0001$], implying that the "learning curve" of the different displays was different.

Table 6.9 Minimum_Criteria Response Time Tests of Within-Subjects Effects

	Source	DF	ANOVA SS	Mean Square	F Value	Pr > F	Sig
Effect	Display	2	6767534790	3383767395	29.86	<.0001	*
Error	Subject*Display	46	5212332183	113311569			
Effect	Trial	11	8748503790	795318526	13.27	<.0001	*
Error	Subject*Trial	253	15160538178	59923076			
Effect	Display*Trial	22	3180072529	144548751	2.80	<.0001	*
Error	Subject*Display*Trial	506	26129051706	51638442			

Examining the *post-hoc* Minimum_Criteria Tukey Response Time grouping by Display Type (Table 6.10) shows the Sprocket display is considerably faster than the

other two displays. This suggests that the Sprocket display is cognitively easier to extract

the data.

Table 6.10 Minimum_Criteria Tukey Response Time grouping by Display Type (Means with the same letter are not significantly different)

Tukey Grouping	Mean	N	Display
A	11784.6	288	Bar Chart
B	7102.4	288	Text Table
C	5107.0	288	Sprocket

Examining the *post-hoc* Minimum_Criteria Tukey Response Time grouping by

Image Presentation Order (Table 6.11), there appears to be a learning curve irrespective

of display.   The first image takes considerably longer to process than any of the

succeeding images. This is probably due to the subject trying to interpret all of the data

on the screen without narrowing their attention to the important features. By the time the

fourth image is displayed, the relevance of specific areas within the display rose in

importance and the operator's attention was drawn to those particular areas.

Table 6.11 Minimum_Criteria Tukey Response Time grouping by Image Presentation Order (Means with the same letter are not significantly different)

Tukey Grouping	Mean	N	Trial
A	16774	72	1
B	11157	72	2
BC	9971	72	3
BCD	8099	72	4
BCD	7660	72	5
BCD	7274	72	6
CD	6619	72	11
CD	6259	72	9
D	6037	72	8
D	5907	72	7
D	5383	72	10
D	4838	72	12

6.1.1.2.3 Response Time for Drop_Dimension

The third question was an exploration into the operator's ability to predict the

future by mentally dropping a dimension of the FoMs and then predicting which mission

140

alternative was best. This was to mimic a specific operation scenario: The case in which current mission computer information has not kept up with the real world. For example, before the mission started, the operator might receive information giving specific information that is not included in the mission displays. The ability of the operator to mentally visualize the changes to the mission alternatives might be important during a mission. This was the most cognitively difficult question, meant to explore situation awareness Level 3 SA.

Unfortunately, display type was **not** found to have a significant effect on predicting the best mission alternative. Although this is a disappointing result, the fact that Level 3 SA errors only make up 3.3% of the total situation awareness errors (see Section 2.1.5.3) alleviates some of the agony. Results from the other three questions (Level 1- and 2-SA) are where the effort and results need to be concentrated.

That being said, there was a significant learning effect from the image presentation order [$F(11,253)=9.37, p<0.0001$] along with a significant interaction between the Display Type and image presentation order [$F(22,506)=1.98, p<0.0001$].

Table 6.12 Drop_Dimension Response Time Tests of Within-Subjects Effects

	Source	DF	ANOVA SS	Mean Square	F Value	Pr > F	Sig
Effect	Display	2	5446634141	2723317070	2.95	0.0626	
Error	Subject*Display	46	42535840506	924692185			
Effect	Trial	11	20747438607	1886130782	9.37	<.0001	*
Error	Subject*Trial	253	50952518401	201393353			
Effect	Display*Trial	22	8351896340	379631652	1.98	<.0001	*
Error	Subject*Display*Trial	506	97059881037	191817947			

Inspecting the *post-hoc* Drop_Dimension Tukey Response Time grouping by Image Presentation Order (Table 6.13)

Table 6.13 Drop_Dimension Tukey Response Time grouping by Image Presentation Order (Means with the same letter are not significantly different)

Tukey Grouping	Mean	N	Trial
A	29617	72	2
AB	28129	72	1
ABC	22428	72	4
BC	21129	72	3
CD	20356	72	9
CD	20082	72	5
CDE	18437	72	8
CDE	18370	72	7
CDE	18170	72	11
CDE	17186	72	6
DE	13517	72	10
E	12006	72	12

6.1.1.2.4 Response Time for Best_Dimension

The final question examines the operator's ability to quickly identify the mission alternative with the best rating for a given FoM. As expected, Table 6.14 shows there was a statistically significant effect of display type on predicting the Best_Dimension [$F(2,46)=8.85$, $p=0.0006$]. The image presentation order also showed a learning curve, i.e., response times tended to improve over time on the same display [$F(11,253)=12.41$, $p<0.0001$]. And finally, there appeared to be an interaction between the display types and the image presentation order [$F(22,506)=8.04$, $p<0.0001$].

Table 6.14 Best_Dimension Response Time Tests of Within-Subjects Effects

	Source	DF	ANOVA SS	Mean Square	F Value	Pr > F	Sig
Effect	Display	2	1724807764	862403882	8.85	0.0006	*
Error	Subject*Display	46	4483541158	97468286			
Effect	Trial	11	5327085375	484280489	12.41	<0.0001	*
Error	Subject*Trial	253	9876518757	39037624			
Effect	Display*Trial	22	6509883486	295903795	8.04	<0.0001	*
Error	Subject*Display*Trial	506	18626469746	36811205			

Looking at the individual *post-hoc* Best_Dimension Tukey Response Time grouping by Display Type (Table 6.15) shows the unexpected result that the Text Table was the fastest response time. This could be because of the presentation of pertinent

information was proximal: once the operator found the correct column label, they did not

have to "search" beyond the column of numbers.

Table 6.15 Best_Dimension Tukey Response Time grouping by Display Type (Means with the same letter are not significantly different)

```
Tukey Grouping    Mean    N     Display
      A          12443.4  288   Bar Chart
      B          10539.8  288   Sprocket
      C           8988.5  288   Text Table
```

Finally, the *post-hoc* Best_Dimension Tukey Response Time grouping by Image

Presentation Order (Table 6.16) showed a learning curve, but one that was more wildly

ordered. One expects the first image presented to have the slowest reaction time, but the

5th image was slowest (14.7 seconds) and the 10th image was in the middle of the pack

(not one of the fastest).

Table 6.16 Best_Dimension Tukey Response Time grouping by Image Presentation Order (Means with the same letter are not significantly different)

Tukey Grouping	Mean	N	Trial
A	14669	72	5
A	14469	72	1
AB	12582	72	2
ABC	12139	72	3
ABCD	11612	72	10
ABCD	11610	72	4
BCDE	10403	72	7
CDE	8922	72	9
DE	8808	72	6
E	7816	72	12
E	7763	72	11
E	7095	72	8

6.1.1.3   Comparison of Response Times

With regards to the first experiment, the learning curve of each question is

examined. The important features to examine are the initial values on the left (this is an

indication of how easily this display type is interpreted by an "inexperienced" operator

for the given question) and the level to which the curve declines (an indication of how easily the display type is interpreted for an "experienced" operator).

For the first question, Figure 6.1 illustrates the learning curves for the respective Display Types. The Bar Chart is the most difficult to interpret for an inexperienced operator, while the sprocket is the easiest. The Bar Chart and the Sprocket appear to level out near the fourth round of questions, but the Sprocket is still slightly lower.
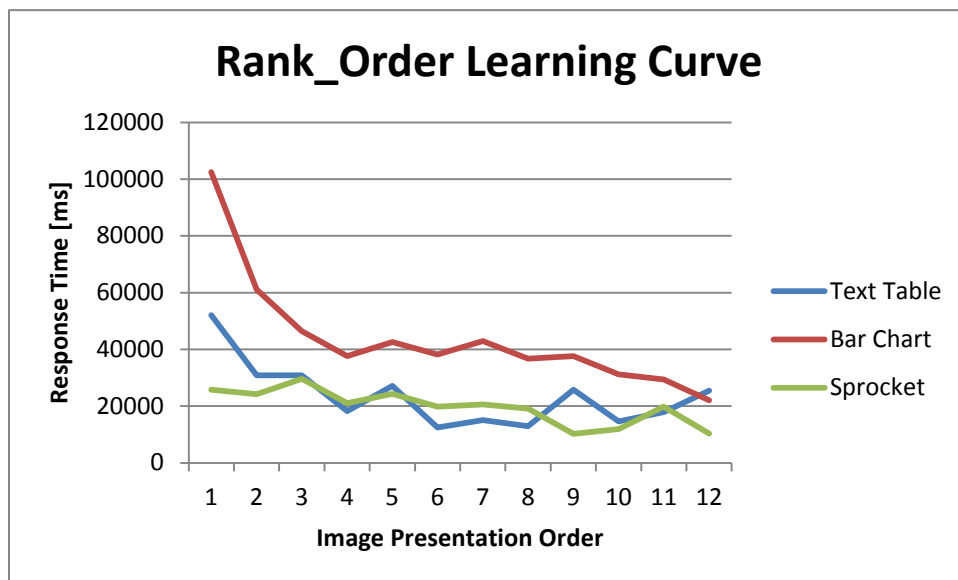
Figure 6.1 Response times for Rank_Order by Display Type

For the second question, the results were similar to the first questions learning curves. However, the Text Table is initially more difficult for the subject to interpret than the first question. The subjects quickly understand what the question is asking, so they are able to anticipate the question and skim more quickly. It should be noted that the Sprocket Display is easily understood by the subjects and has the quickest (or nearly the quickest) response for all questions.
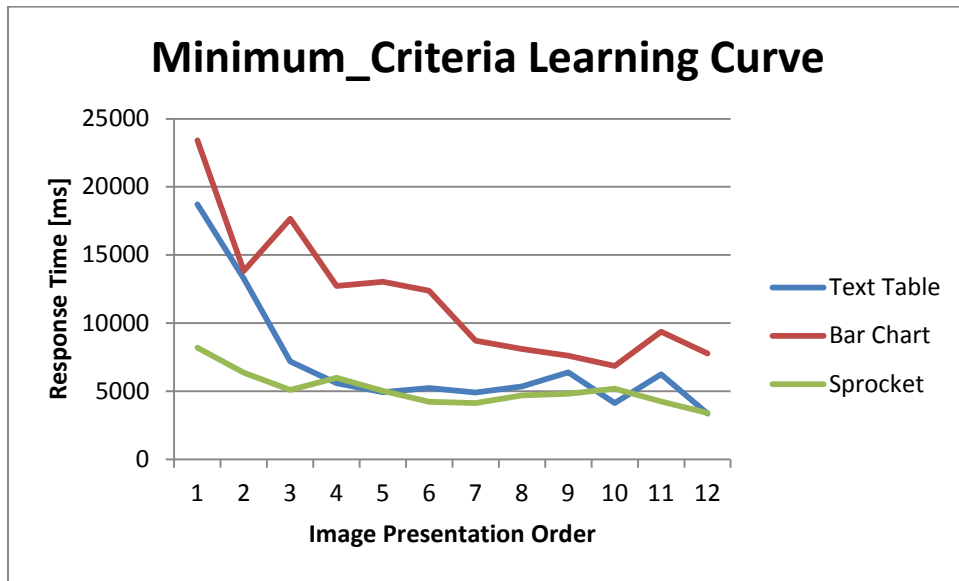
Figure 6.2 Response times for Minimum_Criteria

Dropping a Dimension is significantly more difficult than the previous two questions, requiring the subject to mentally recalculate the "best route" based on the information presented to them.
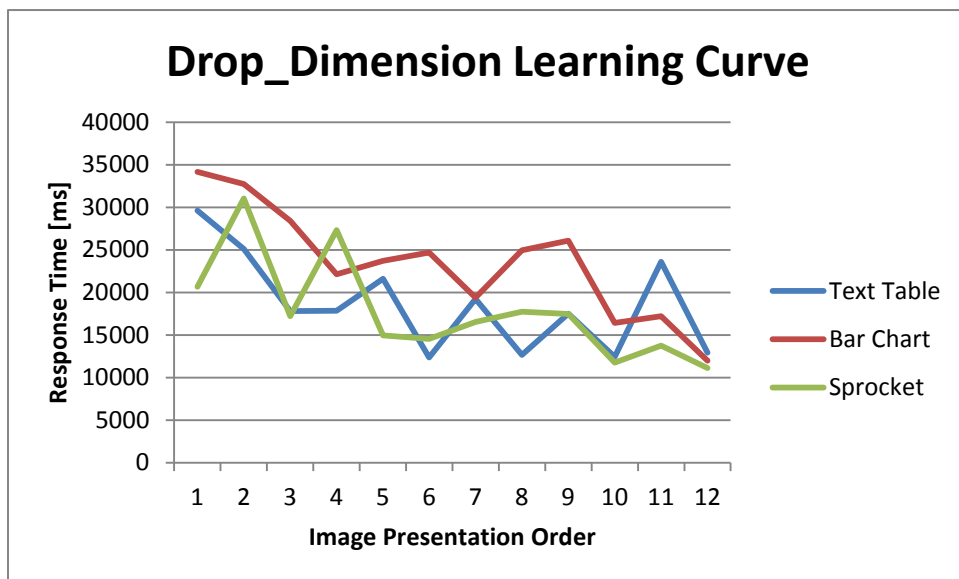
Figure 6.3 Response times for Drop_Dimension

The Text Table was the quickest aid when determining the best dimension. The learning curve is flattest and seems to be the easiest to mentally compute. That said, the Sprocket display is not that far behind.
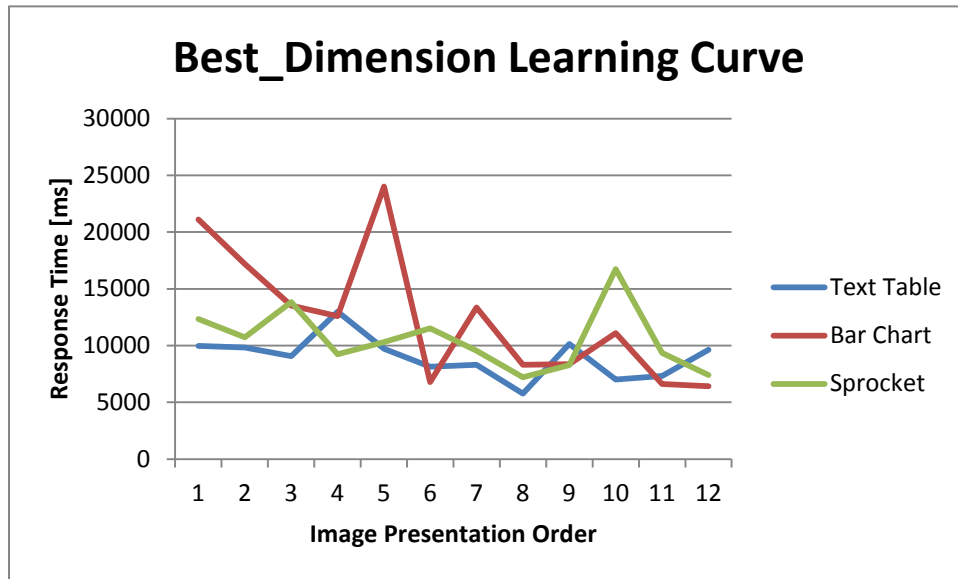
Figure 6.4 Response times for Best_Dimension

Taking all four of the learning curves, the Sprocket display seems to be the easiest to learn for three situations and is near the bottom for the fourth. This implies that the Sprocket display is easiest to learn among these three displays.

6.2 *System Monitoring Experiment*

The data was comprised of answers to questions related to system monitoring of three dynamic displays presented to the subjects. The system monitoring questions were based on system health information – see Table 5.5. The data collected for this experiment used questions to ascertain the ease of use as a system monitoring tool for each display. The text table represents an advancement of the current UAV display technology (circa 2007) used in a Predator Ground Control Station (GCS) Variable

146

Information Tables (VIT). The analog gauge was chosen as an exemplar of a current attempt at a "graphical user interface".

The displays were dynamic animations meant to elicit quickly identified solution patterns for the operator. The questions were designed to explore specific decisions considered typical within a multi-UAV mission.

The overall results expanded in the next section show that the Sprocket and Analog Displays were significantly faster than the Text Display. On the other hand, the Sprocket and Text Displays were significantly more accurate than the Analog Display. This infers that if one wants both speed and accuracy, the Sprocket Display would be preferred over the Analog and Text Displays.

6.2.1 Analysis of Data

In essence, this experiment had interesting results, not all of them expected. The experiment examined a primary task (simulating a surveillance mission by maintaining a target box in the crosshairs), a secondary task (simulating a system monitoring task by responding to gauges that go out of tolerance), and probe questions that examine the operators' situation awareness. The data for the operators' answers to the static displays were analyzed using JMP® 9.0 for windows.

This experiment examined three objectively measureable facets: (1) the root mean square error (RMSE) of the distance from optimal surveillance for the Primary Task; (2) the response to when a gauge went out of tolerance for the Secondary Task; and (3) the response to SA questions presented to the operator approximately every minute.

6.2.1.1 Primary Task Distance from Optimal – Root Mean Square Error

The primary task simulated a UAV mounted camera that constantly moved because of swirling wind, UAV movement and a bad tracking gear. These three character flaws required the operator to constantly monitor and correct the camera direction to maintain the target under the camera crosshairs.

To examine the data for the primary task, the distance from optimal position had to be calculated and corrected. With respect to the primary task, the research examined the root mean square error (RMSE) distance in pixels from the actual position of the target and the optimal position, i.e., being completely within the crosshair scope. Since the crosshair was larger than the target, the optimal position, i.e., target being fully within the crosshairs, was a ± 10 pixel range. That is, if the target is located at $x=20$ and $y=-5$, to calculate the Euclidean distance, one had to first map the $<x,y> \rightarrow <\dot{x},\dot{y}>$ to incorporate the ± 10 pixel range: $\dot{x}=floor(abs(x-x_0))=floor(abs(20-10),0)=10$ and $\dot{y}=floor(abs(y-y_0))=floor(abs((-5)-10),0)=0$. The final Euclidean distance uses these corrected values $<\dot{x},\dot{y}>$; $RMSE_{<x,y>} = \sqrt{\dot{x}^2 + \dot{y}^2} = \sqrt{10^2 + 0^2} = 10$.

With respect to the correcting the data sampling frequency, each 25 minute experiment run could generate more than 2500 data points. The first attempt in processing these all of these data points created an analysis that was overwhelmed by the denominator (used in the statistics), since the variation between each pair of points was relatively smooth. That is, the total variability was too small over such a large number of readings. The second attempt was to perform a running average of the data to create a

lower statistical denominator. This lowered the statistical denominator; but as a side effect, the data was smoothed and again the variability was not evident. There were two versions of this smoothing: (1) averaging every 100 points irrespective of length of run time, $p' = \sum_{n=1}^{100} p_n \Big/ 100$; and (2) averaging every minute irrespective of number of data points to be averaged, $p' = \sum_{n=1}^{N} p_n \Big/ N$, where $N = count(t_i - t_{i-1})$.

The third attempt used the data point at the 60 second mark, essentially allowing the denominator to be lowered while at the same time allowing the data to remain unsmoothed. The obvious side effect of ignoring most of the raw data is deemed acceptable for the following reason. Consider the raw data to be equivalent to looking at an object at very high resolution, such as a painting of a horse by Remington, a famous American Painter of southwest life in the 1800s. However, a dot-to-dot drawing of a horse by a three year old toddler is a simpler representation. Both images reveal a pattern that can be interpreted as "horse", but the simpler dot-to-dot image emphasizes bold features, while Remington's picture is appreciated for its fine detail. This simplification is equivalent to choosing a data point at each minute. The underlying "tracking" pattern might emerge. In other words, the coarse pattern ignores the jittery intermediate noise to focus on the gross pattern.

Table 6.17 shows the statistical results of these snapshots. There were some significant differences among the users, displays and their interactions. This necessitates an investigation of the underlying causes. The fact that the subjects are significantly different is problematic.

Table 6.17 Primary Task Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Display	2	2	39255.99	3.1193	0.0445*
Subject*Display	32	32	422320.08	2.0974	0.0004*

Post hoc analysis of the RMSE distance among the three displays shows that the Sprocket is significantly faster than the Text Table, but it is indistinguishable from the Analog Display. Also, that the Text table is indistinguishable from the Analog Display. One way of interpreting this is that the subject has less mental workload on the secondary task, thus making the primary task easier to perform.

LS Means Plot

LSMeans Differences Tukey HSD

α= 0.050   Q= 2.3466

Level		Least Sq Mean
Text Table	A	52.794138
Analog Gauge	A B	47.936642
Sprocket	B	39.300278

Levels not connected by same letter are significantly different.
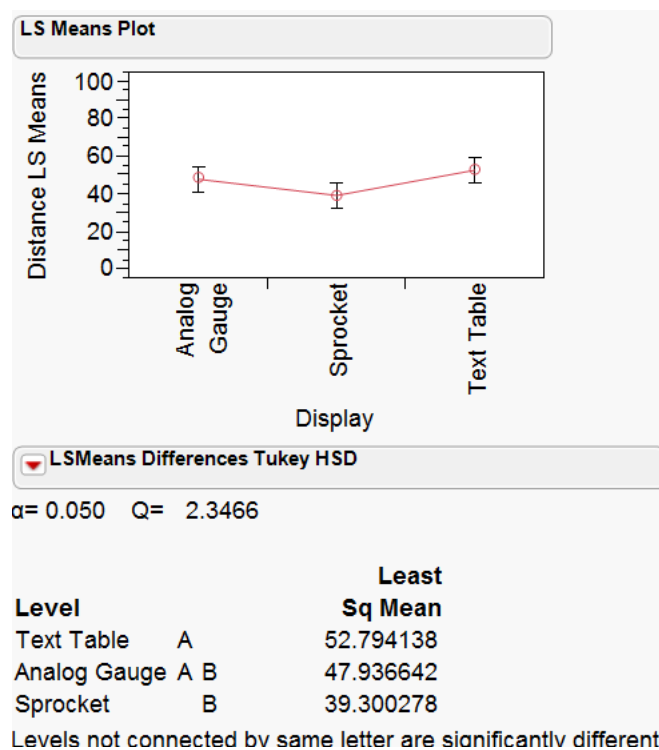
Figure 6.5 Post hoc analysis of the RMSE distance for the three displays

6.2.1.1.1 Primary Task Distance from Optimal (RMSE) with respect to Subject

Examining the *post-hoc* subjects Response Time grouping results for multiple comparisons with an alpha level set to α=0.05 shows that Subject 17 was unusually slow compared to the rest of the subjects (Figure 6.6). That subject was "asleep with his eyes open", not reacting to the gauges.

Removing the 17th subject from the data because of its outlier status leaves a different, more interesting picture. Without the anomalous subject, we still have statistically significant differences for the model [$F(2,34)=2.1503$, $p=0.0002$].

LS Means Plot

LSMeans Differences Tukey HSD

α= 0.050   Q= 3.49594

Level		Least Sq Mean
17	A	820.16063
12	B	147.45265
7	B	118.28266
8	B	92.28644
13	B	69.21138
14	B	57.87491
10	B	43.90806
6	B	41.82256
15	B	33.58365
1	B	26.97862
4	B	26.90348
5	B	23.97791
3	B	22.52294
9	B	21.72541
16	B	19.68806
11	B	16.76750
2	B	16.09021
18	B	14.43289

Levels not connected by same letter are significantly different.
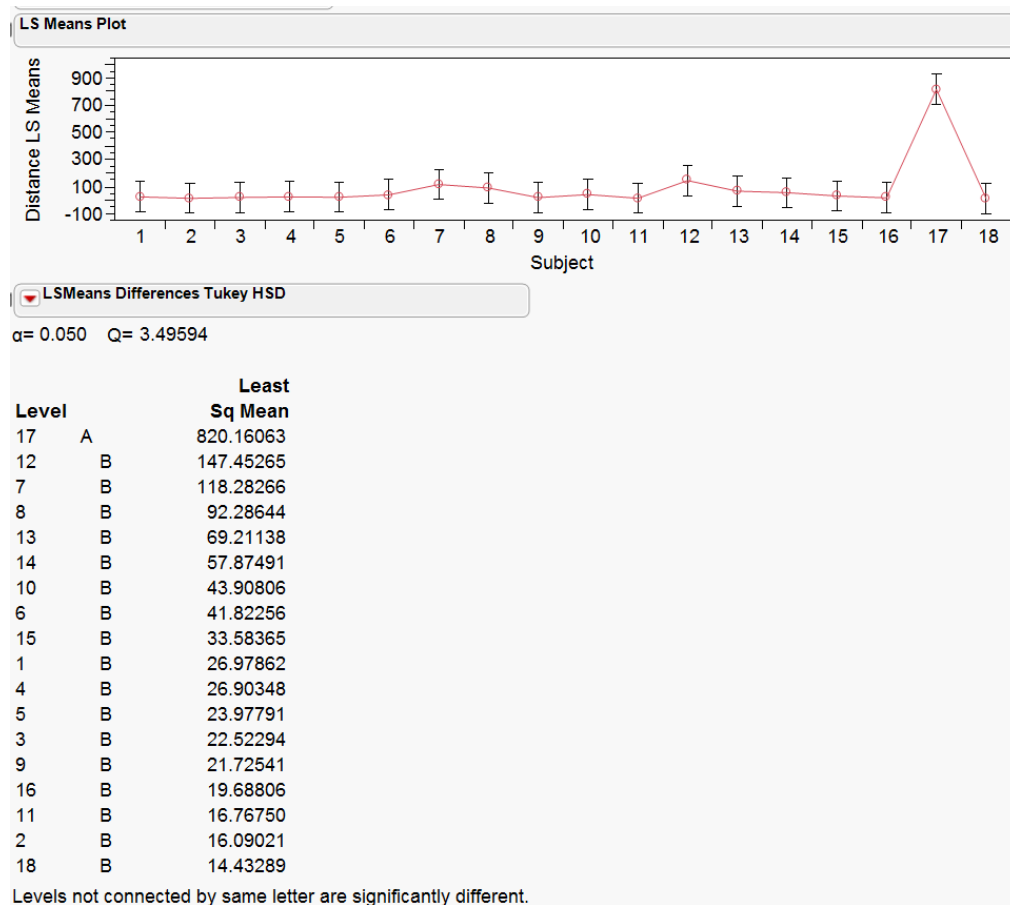
Figure 6.6 Users Least Squares Plot showing subject 17 being extremely slow.

Furthermore, Subject 17's anomalous behavior was so abhorrent that all other differences among the subjects were drowned out into a wash of similarity (Figure 6.6). Now examining the Effect Tests shows that the Display are statistically different [$F(2,32)=3.1193$, $p=0.0445$]. Without Subject 17, there is appears to be much more variability within the remaining subjects (Figure 6.7). However, one should note that the y-axes are two drastically different scales in both figures. The variability would not be as obvious if the scales were the same. The subjects with the highest values (Level A) probably quickly decided to assign Task 2 (system monitoring) a higher priority than Task 1 (Surveillance). The subjects with the lowest values (Level F) prioritized the Surveillance as a higher priority than the system monitoring. Those subjects in between either had a more balanced approach or changed priorities over time.

LS Means Plot

LSMeans Differences Tukey HSD

α= 0.050   Q= 3.46592

Level						Least Sq Mean
12	A					147.45265
7	A B					118.28266
8		B C				92.28644
13			C D			69.21138
14			C D E			57.87491
10				D E F		43.90806
6				D E F		41.82256
15				D E F		33.58365
1					E F	26.97862
4					E F	26.90348
5					E F	23.97791
3					E F	22.52294
9					E F	21.72541
16					E F	19.68806
11					F	16.76750
2					F	16.09021
18					F	14.43289

Levels not connected by same letter are significantly different.
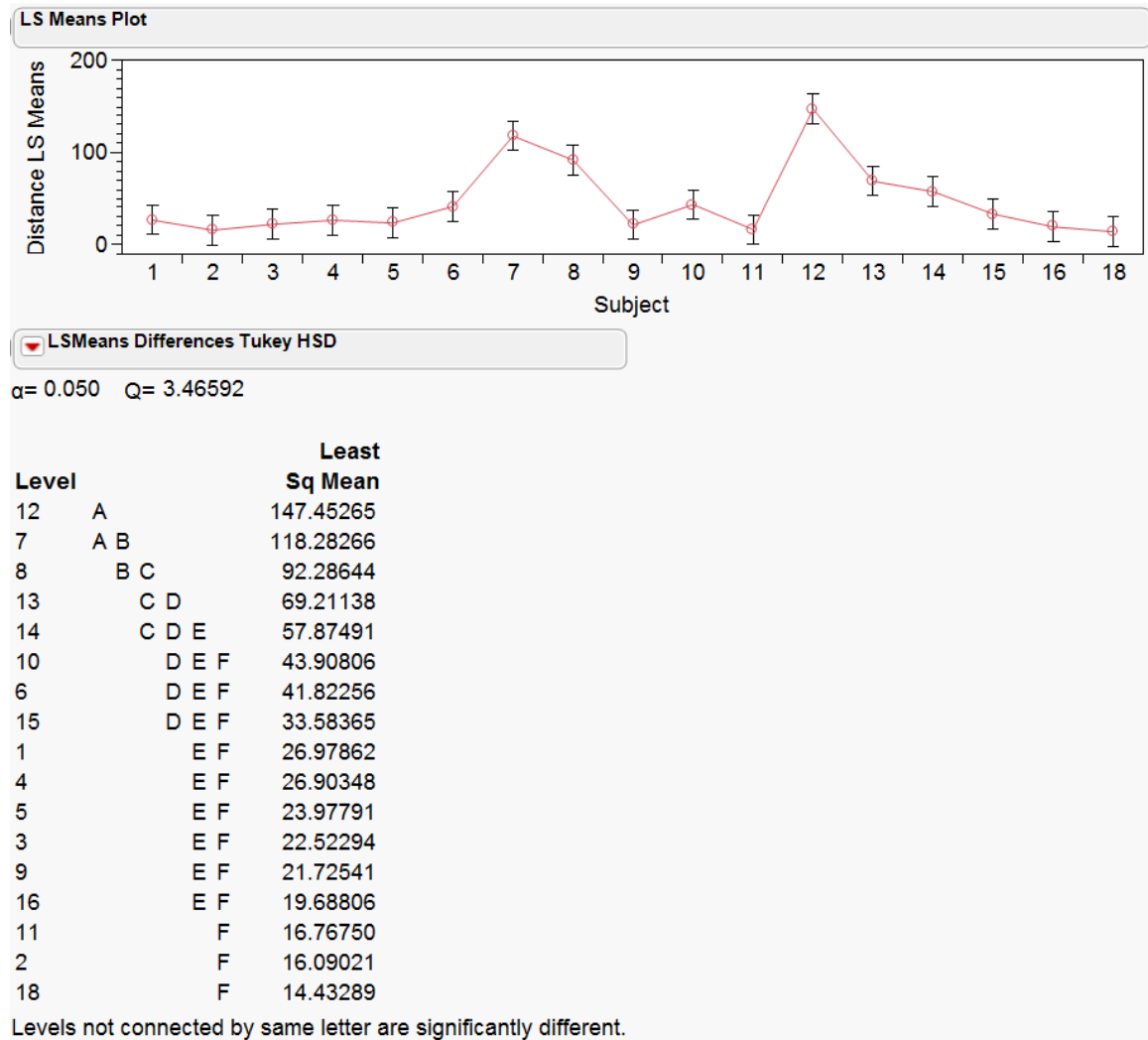
Figure 6.7 Least Squares Means Plot and LSMeans Differences Tukey HSD with Subject 17 removed as an outlier

Examining the distance over time with respect to subjects shows there are no significant changes over time (Table 6.18). This implies that there was no changing of priorities over time and any strategy with which a subject started was the same strategy with which they ended the experiment.

Table 6.18 Effects Tests of Distance over Time with respect to Subject

Effect Tests					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Subject	16	16	99317.924	1.2121	0.2508
Minute	1	1	45.133	0.0088	0.9252
Subject*Minute	16	16	83280.223	1.0164	0.4358

Just to complete this reflection on changing strategies, we examine the distance over time with respect to the displays (Table 6.19). For a second time, there are no apparent differences among the displays as time progressed. So once again we are left with the conclusion that whatever strategy a subject began the experiment with was the same one that they ended with.

Table 6.19 Effects Test of Distance over Time with respect to Display

Effect Tests					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Display	2	2	28231.486	2.1910	0.1122
Minute	1	1	580.003	0.0900	0.7642
Display*Minute	2	2	14230.429	1.1044	0.3317

6.2.1.1.2 Primary Task Distance from Optimal (RMSE) with respect to Display

From Table 6.17, we know that the displays are significantly different. Examining the least squares means plot (Figure 6.8) confirms that there is a difference between the Sprocket and Text Table, but no statistic difference between Sprocket and Gauge and Analog Gauge and Text Table. The Sprocket is much better than the current Text Table UI technology.

LS Means Plot

LSMeans Differences Tukey HSD

α= 0.050   Q= 2.3466

Level		Least Sq Mean
Text Table	A	52.794138
Analog Gauge	A B	47.936642
Sprocket	B	39.300278

Levels not connected by same letter are significantly different.
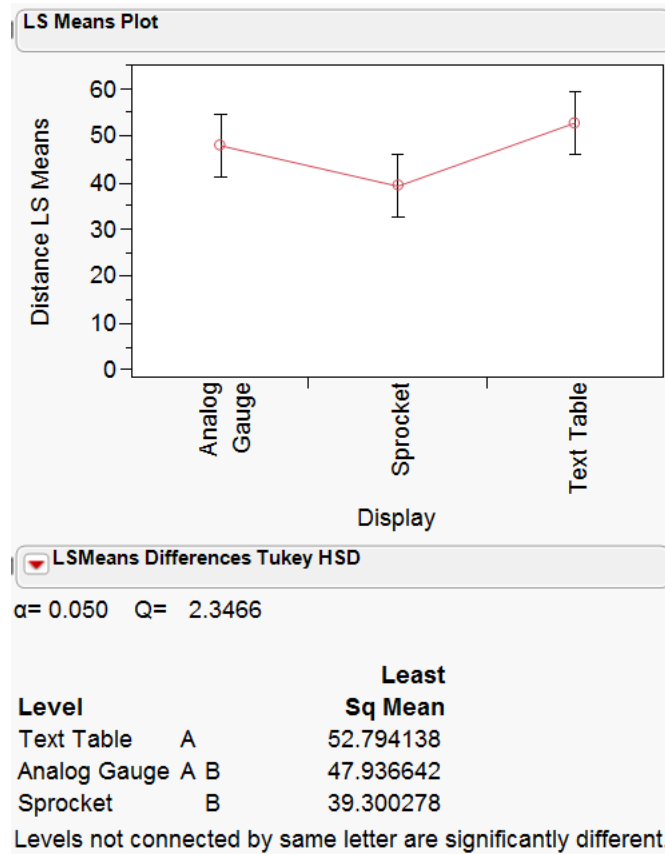
Figure 6.8 Least Square Means of Primary Task RMSE with respect to Display

6.2.1.2   Secondary Task: Response to when a Gauge goes Out-of-Tolerance

With respect to secondary task of system monitoring, the research examined the speed and accuracy of response. The response time was the difference in time between the gauge going out of tolerance and the subject correcting the condition by clicking on the gauge. Clicking on the display reset the gauge to initial conditions, and the scripted animation would restart. Accuracy of response considered whether the user correctly responded to an out of tolerance condition or incorrectly clicked on a gauge (error). The two types of errors a subject could have: (1) click on gauges that were not out of tolerance (false click); or (2) click on the same gauge twice (double click).

155

6.2.1.2.1 Secondary Task Response Time

There was a statistical difference for the speed the operator responded found among the users [$F(50,5774)=569.5$, $p=0.0001$]. And indeed, examining the effects tests confirms that there is a significant difference among users, blocks and the interaction between users and blocks (Figure 6.9).

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
User	16	16	9393379	910.2113	<.0001*
Block	2	2	4052901	3141.784	<.0001*
User*Block	32	32	13930620	674.9332	<.0001*

Figure 6.9 Effects tests of Secondary Task Response Time

This finding appears to be very good, so we delve further into the effects (Figure 6.10). After examining the Least Means Square plot for the users, one finds that there are a couple of peaks in an otherwise flat response time. And the Least Means Square Differences Tukey HSD for the users confirms there are four levels of performance.

LS Means Plot

α= 0.050   Q= 3.45999

Level				Least Sq Mean
14	A			256.26595
8		B		64.99392
13			C	38.87440
7				D	6.21670
12				D	4.59297
10				D	4.47334
6				D	4.25891
11				D	4.16944
18				D	3.51344
16				D	3.42194
15				D	3.14953
9				D	2.87195
3				D	2.85895
4				D	2.61156
5				D	2.37530
2				D	1.92293
1				D	1.45387

Levels not connected by same letter are significantly different.
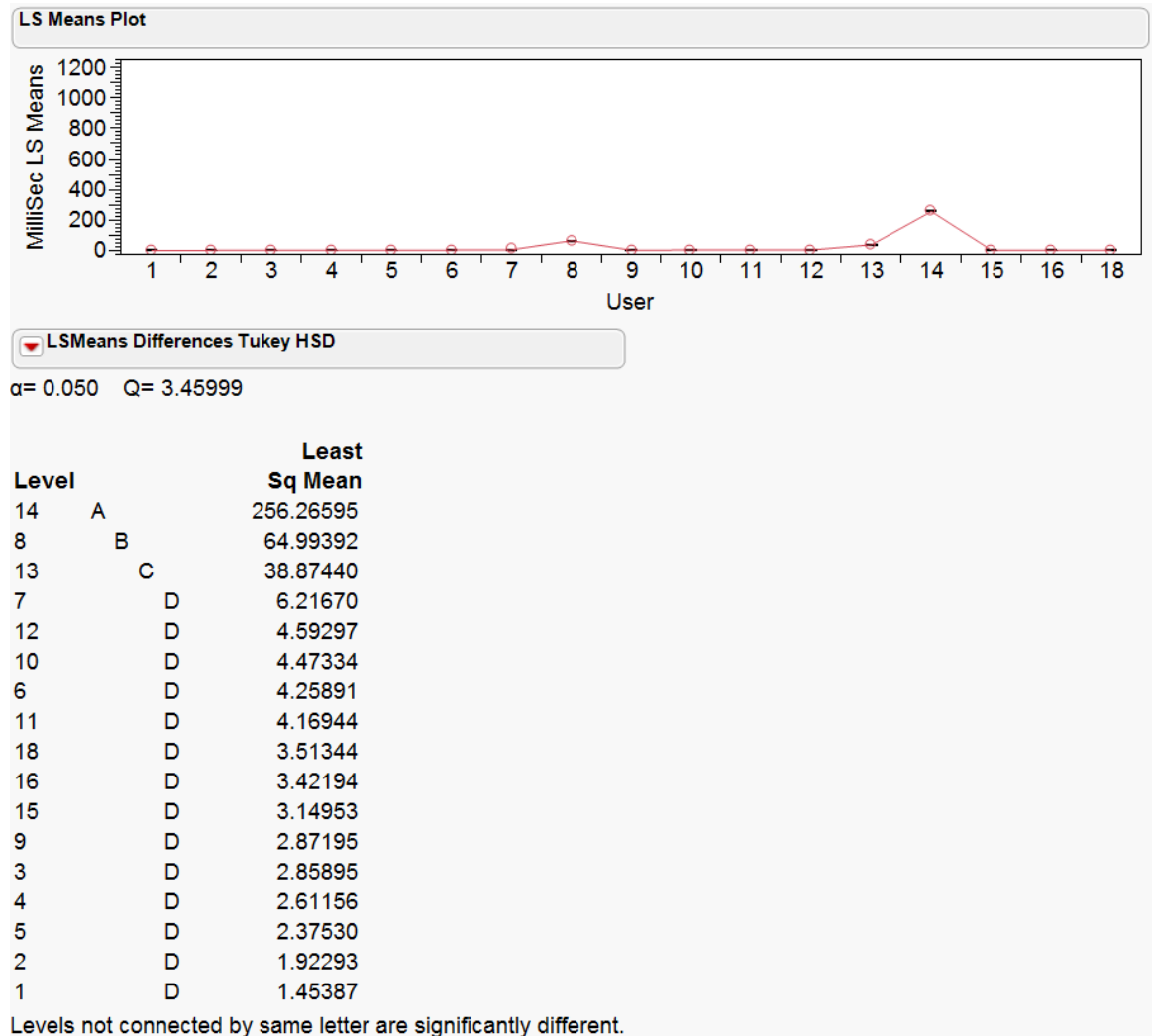
Figure 6.10 Least Square Means of Secondary Task effects

Going back to the raw data finds some obvious strategies employed by four subjects; namely giving very high priority to the primary task to the exclusion of the secondary task. To verify this conclusion, a simple count of errors committed per subject was tabulated and the results were astonishing (Figure 6.11). The top row, Missing, shows how many responses the subject had without committing an error – typically ranging between 300 and 400. On the other hand, the row labeled "1" represents the number of errors committed if there were more than 100 committed. For subjects 6, 8, 13

and 14, there appears to be a problem. Their error count was a significant proportion to their total responses. The raw data shows that these four users reset all gauges when one went out of tolerance so that they could concentrate on the primary task. Subject 14 hit upon this strategy early, hence his low count for non-errors and his very high count for errors. Subject 8 & 13 found this strategy after completing at least one run using the "correct" strategies suggested in the preliminary instructions given to all subjects. Finally, Subject 6 performed most of the tasks within parameters, but made an excessive number of double clicks, *i.e.*, he did not always wait for the first click to register with the Windows® operating system, notoriously known for not being real-time computing.

Error	User																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	18
Missing	397	373	375	381	367	308	359	277	376	296	369	360	297	159	377	379	375
1	0	0	0	0	0	133	0	853	0	0	0	0	213	656	0	0	0

Figure 6.11 Count of Errors for each subject in secondary task

Looking at the same count by display shows that there does not appear to be any evidence that displays were immune to errors, but the Sprocket had the fewest. Furthermore, the one error that occurred more frequently with the Sprocket was the "nearest neighbor" error. In this error, the neighboring FoM pie pieces were clicked. However, there were only 12 of these errors, which is an insignificant number when compared to the total events (25,000+). This was probably because of the button's design, and can probably be programmatically corrected by disallowing clicks near the center of the pie piece.

Error	Block		
	Analog	Sprocket	Text
Missing	2271	2107	1447
1	661	491	703

Figure 6.12 Count of Errors for each display in secondary task

This discussion is to address the possibility of dropping the subjects or displays that appear to be anomalous. Obviously that displays are not going to be the deciding factor here, all three displays seem reasonably similar. Examining the raw data suggests that Subject 6 was not "redefining" the parameters, but was just eager. So, running the same tests with Subjects 8, 13, and 14 excluded finds that all of the remaining users are indeed similar, but there is a significant difference among the displays (Figure 6.13).

Fixed Effect Tests

Source	Nparm	DF	DFDen	F Ratio	Prob > F
User	13	13	25.73	1.4401	0.2078
Block	2	2	25.77	67.1312	<.0001*

Figure 6.13 Fixed Effects Tests without Subject 17 (previous sections) and Subjects 8, 13, and 14

Knowing that there is a difference in the Blocks (Displays), we examine those effects the displays had on the model using the Least Squares Means plot and the Least Squares Means Tukey HSD. From these we conclude that there is a significant difference in the response speed of the displays – the Text Table is much slower.

LS Means Plot

LSMeans Differences Tukey HSD

α= 0.050

Level		Least Sq Mean
Text	A	7.9045250
Analog	B	1.7224631
Sprocket	B	0.6161930

Levels not connected by same letter are significantly different.
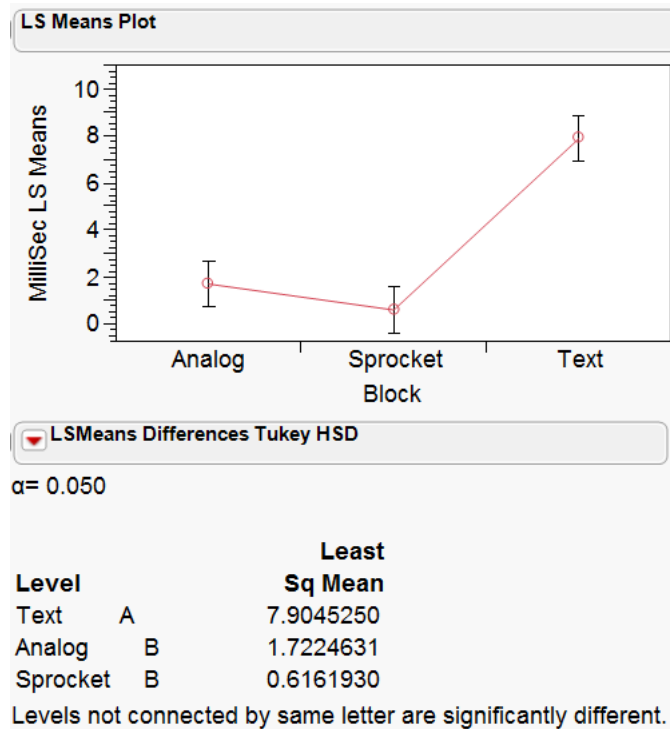
Figure 6.14 Least Square Means of Secondary Task Display effects

6.2.1.2.2 Secondary Task Accuracy

When it comes to the accuracy of responding to the out of tolerance there were two error conditions examined: double clicking and false clicking. Double clicking typically arises from the subject observing an out of tolerance event and clicking on the target gauge twice. This can normally be explained be two conditions: (1) the subject double clicked because of the lag time inherent in the Java®/Windows® platforms. There were no significant differences among the displays for all Errors [$F(2, 39)=2.7503$, $p<0.0750$], just the Double Clicking [$F(2, 39)=0.4849$, $p<0.6194$], nor False Clicking [$F(2, 39)=2.8819$, $p<0.0680$].

That does not mean that the errors did not affect the experiment. The number of user generated events is the sum of the correct responses and the incorrect responses

160

(Errors). In general, if there were no errors, then all to the events for the displays should have balanced out, because of Latin Squares balancing the displays and scripts. However, this was not the case. In fact, because of the additional clicking done for the errors, the Analysis of Variance for the entire model revealed that there was a difference among the displays [$F(2, 39)=39.3629$, $p<0.0001$].

The Analog Gauge had significantly higher subject initiated events than the Text Table (Table 6.14). This means that the subjects were "busier" because they were making errors for themselves. The significance of this is that the more errors created, the busier the user – in interesting paradox to say the least.

LS Means Plot

LSMeans Differences Tukey HSD

$\alpha= 0.050$

Level		Least Sq Mean
Text	A	727.57143
Analog	B	240.85714
Sprocket	B	78.35714

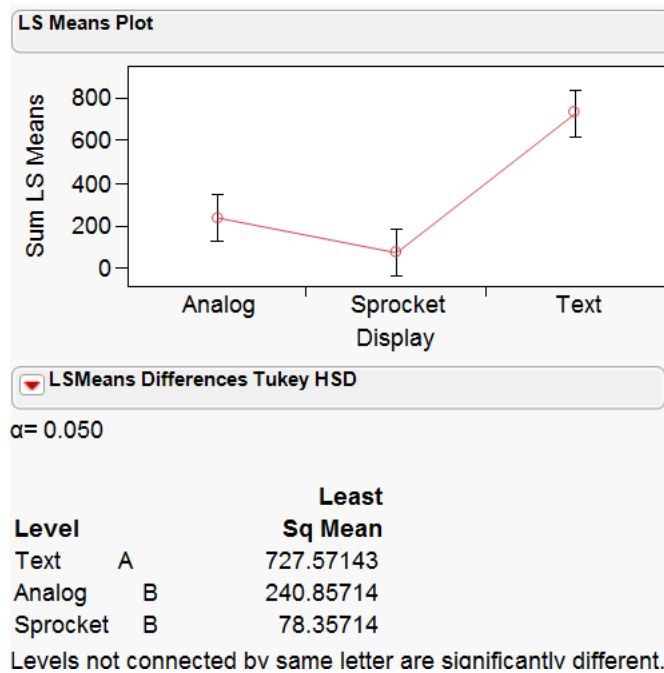Levels not connected by same letter are significantly different.

Figure 6.15 LS Means Plot and LS Means Diferrences Tukey HSD for the count of Subject Initiated Events: correct response + error response

In addition to the count, the amount of time wasted to respond for these errors has a significant difference among the displays, [$F(2, 39)=19.0048$, $p<0.0001$]. This is the

total amount of time that is wasted for all of the errors for a given display (Table 6.15). In other words, if the time between the two clicks of a double click takes one second and there were fourteen Double Click errors for the Analog Gauge display, then fourteen seconds were wasted during this experimental run. The amount of time wasted performing errors is significantly less for the Text Table, while the Analog Gauge has significantly more time wasted. This implies that when an error was made in the Text Table, they were of short duration, while the Analog Gauge errors were of relatively long duration.

LS Means Plot

LSMeans Differences Tukey HSD

α= 0.050

Level		Least Sq Mean
Analog	A	146.00000
Sprocket	B	133.85714
Text	C	118.57143

Levels not connected by same letter are significantly different.

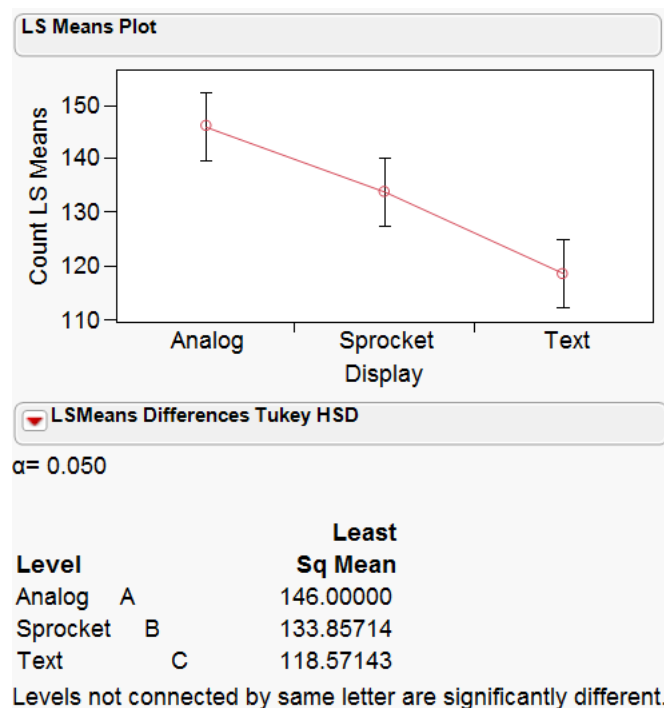**Figure 6.16 LS Means Plot and LS Means Differences Tukey HSD for the time to respond of Subject Initiated Events: correct response + error response**

6.3 *Full System Experiment: Displays within a full simulator*

This experiment was performed at Wright Patterson AFB with active duty and retired military UAV operators. The retired UAV operators were currently working as

subcontractors, training new UAV operators in Victor Valley, CA. The pilot experience was 14.7±8.2 years, while the pure UAV operator experience was 8.3±4.5 years.

The software used for this experiment, MAGE, is not the official USAF simulator software. The MAGE software is a testbed to examine next generation user interfaces in support of command and control of multiple UAV missions. As such, the subjects were not familiar with the operation of the software and received a 45 minute tutorial and sample mission to familiarize them with MAGE.

Within the MAGE software were multiple new user interfaces, the Visual Thinking Sprocket being just one. Others user interfaces included a temporal planner, map overlays, mission timeline monitor and a planning "sandbox". As mentioned earlier, the subjects were instructed during training to "Think Out Loud", verbalizing their impressions of the system during the scenario, so that we could have some insight into their thinking. One of the subjects, upon seeing the Visual Thinking Sprocket for the first time, said "Wow!" Asked to elaborate, he stated that it was a "confusing looking" display. However, the third time that the Visual Thinking Sprocket was presented to him in the simulation he said "Oh, I get it".

The subjects were specifically asked to subjectively rate the workload of the different decision aids used within MAGE using NASA TLX. For task performance, significant differences were observed over task periods ($F(2,39) = 15.1133$, $p < .0001$). Examining the LS Means Plot and Differences Tukey HSD indicates that there is a significant difference among the displays, with the Sprocket being significantly better than the other displays.
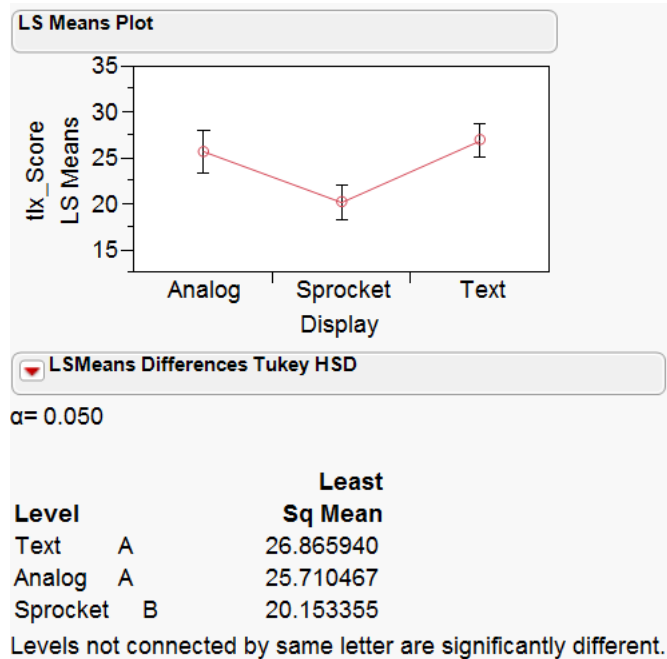
Figure 6.17 Mean Values of NASA-TLX scores for the Text, Analog and Sprocket displays

Post mission comments from the subjects indicated that the sprocket was "strange at first, but simple to use", "analog display was difficult to read – I had to squint", and "analog and text displays take up too much screen space". These comments probably were a major effect on the eventual ranking of the displays.

6.4 *Conclusions from experiments*

From the Decision Support Experiment, the Sprocket design was found to be a faster and more accurate display as a decision aid in general. Furthermore, the experimental results show that the Sprocket display is generally easier to learn than either of the other two displays. Looking back at the specific research questions originally asked in Section 4.1., the Decision Support Experiment findings are summarized in Table 6.20.

Table 6.20 Research questions for Decision Support Experiment: Legend – Visual Thinking Sprocket (VTS), Bar Chart (BC) and Text Table (TT)

	Research Question	**Findings**
Accuracy 1	Is there a significant difference in accuracy to obtain good rank order solutions using cognitively sensitve supervisory control of multiple UAVs?	YES: There is a significant difference – the VTS was significantly more accurate than BC or TT. BC and TT were similar. [VTS>BC=TT]
Accuracy 2	Is there a significant difference in accuracy to determine global status information using cognitively sensitive supervisory control?	YES: There is a significant difference – the VTS and TT were significantly more accurate than BC. VTS and TT were similar. [VTS=TT>BC]
Accuracy 3	Is there a significant difference in accuracy to envision new solutions using mental visualization?	YES: There is a significant difference – the VTS was significantly more accurate than BC or TT. BC and TT were similar. [VTS>BC=TT]
Accuracy 4	Is there a significant difference in accuracy to determine local status along a specific dimension using cognitively sensitive supervisory control?	NO
Response Time 1	Is there a significant difference in time to obtain good rank order solutions using cognitively sensitve supervisory control of multiple UAVs?	YES: There is a significant difference – the VTS was significantly faster than TT and TT was significantly faster than BC. [VTS<TT<BC]
Response Time 2	Is there a significant difference in time to determine global status information using cognitively sensitive supervisory control?	YES: There is a significant difference – the VTS was significantly faster than TT and TT was significantly faster than BC. [VTS<TT<BC]
Response Time 3	Is there a significant difference in time to envision new solutions using mental visualization?	**NO: There was no significant difference among the displays**
Response Time 4	Is there a significant difference in time to determine local status along a specific dimension using cognitively sensitive supervisory control?	YES: There is a significant difference – the TT was significantly faster than VTS and VTS was significantly faster than BC. [TT<VTS<BC]

The last line of each finding represents the ordering and relative statistical findings, where the first display has the "best" mean value, and the operators < and > refer to statistically significant differences, while the operator = means statistically equivalent. For example, Accuracy ones findings, [VTS>BC=TT], can be interpreted as "the VTS was significantly more accurate than BC or TT. BC and TT were similar".

The System Monitoring Experiment found the VTS is either the best or statistically indistinguishable from the best display for speed and correct responses. However, the experiment found TT had significantly fewer errors than the VTS, which had significantly fewer errors than the AG. The color coding to the TT was probably a large factor in this accuracy. The System Monitoring Experiment findings are summarized in Table 6.21.

Table 6.21 Research questions for Experiment 2: Legend – Visual Thinking Sprocket (VTS), Analog Gauge (AG) and Text Table (TT)

Experiment	Research Question	Findings
Distracter: Root Mean Square Error [RMSE]	Can the operator perform the primary task (keeping the target centered) better using cognitively sensitve human supervisory control of multiple UAVs? [RMSE]?	Yes: The VTS and AG were significantly more accurate than the TT. There was no significant difference in accuracy between the VTS and AG. [VTS=AG>TT]
Monitor Task: Accuracy	Is there a significant difference in accuracy when monitoring out of tolerance values?	Yes: The VTS and AG were significantly more accurate than the TT. There was no significant difference between the VTS and AG. [VTS=AG>TT]
Monitor Task: False Positive	Is there a significant difference in false positives when monitoring out-of-tolerance values?	Yes: The TT were fewer errors than the VTS which in turn had fewer errors than the AG. [TT>VTS>AG]
Response Time	Is there a significant difference in time to respond to out-of-tolerance monitored values using cognitively sensitive supervisory control?	Yes: The VTS and AG were significantly faster than the TT. There was no significant difference between the VTS and AG. [(VTS=AG)>(AG=TT)]

The final experiment used subject matter experts that were quite positive about the Sprocket display. Despite the small pool of subjects, the fact that they were experts and not enlisted college students reflects positively on the conclusions. Here, the question asked and answered was that the SME preferred all three modified displays over

166

the current VIT data collection system. However, they considered the Visual Thinking

Sprocket to be significantly easier to use than the Text Table or the Analog Gauge.

7 Research Contributions and Significance

Technological advances in propulsion, sensors, wireless communication, and other areas allow us to have quite sophisticated UAVs. These advances have overwhelmed the UAV operators/supervisors with data and information. This research presented a new model of human cognition that attempts to implement a cognitively sensitive human-computer interface. The interface presented information to the operator/supervisor in a graphical format that more closely follows a visual thinking paradigm.

Perhaps the most important contribution is a simple three step practical design loop for the implementation of UIs, which was in turn used to design and assess the displays used in this experiment (Table 7.1).
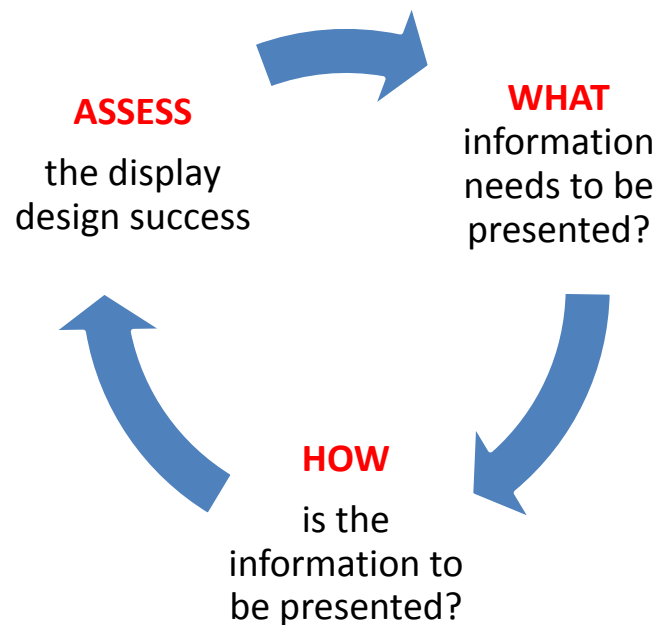
ASSESS
the display
design success

WHAT
information
needs to be
presented?

HOW
is the
information to
be presented?

Figure 7.1 Iterative process to designing a display (copied from Figure 1.1)

This simple, yet elegant framework can be used by a researcher or programmer to guide the steps of developing a UI that is cognitively sensitive. Just like McNeese's research framework guides the researcher through the necessary steps but does not dictate the tools to be used, this framework guides the designer through the necessary steps (what, how and assess), and the underlying tools are those available to the designer (user modeling, NDM, semiotic analysis, VT, SA, etc.) This framework encourages an iterative spiral design methodology that is easy to milestone.

Secondly, the use of visual thinking features when designing a display was shown to be a viable option. The Sprocket design was found to be either best or tied for best in almost all statistical measurements. Because the Visual Thinking Sprocket is cognitively sensitive, the operator was able to identify patterns quickly and efficiently, especially obvious in the decision support experiment. The interpretation of difficult data was not mentally taxing, but was grasped quickly.

This research showed that the Sprocket was effective in the UAV domain, and by extension it can be reasonably assumed that the sprocket may be effective in the decision aid and system monitoring family of problems.

The Operator Function Model was found useful helping to develop the Semiotic Analysis in a systematic manner that was computationally implemented and evaluated. This means that development of operator models can help in semiotic analysis, because Semiotic Analysis suggests that signs should be put into cultural reference. In this case, the culture is UAV operators. Understanding the culture through the OFM assisted in applying that semiotic analysis to the UAV operator culture.

Furthermore, one of the side effects of the Naturalistic Decision Making interviews was designed extensibility in the Visual Thinking Sprocket. Because of the UAV culture, different levels of bureaucracy had different requirements. For example, the Secretary of the Air Force may mandate certain Figures of Merit (FoM), such as a calculated Probability of Success. At the same time, the UAV constructor might require a FoM that indicates the wind speed because of possible stall conditions. Similarly, the wing commander might require a specific FoM and the operator might desire yet another FoM. The higher levels of bureaucracy FoMs could not be altered by the lower levels. The Visual Thinking Sprocket is designed to be flexible, allowing between five and twenty FoMs. Besides adding mandatory FoMs, each level of bureaucracy can also affect the assigned weights.

Besides this reason for the flexibility, the domain is complex and evolving. FoMs that were considered important previously might not be important now. For example, a No Fly Zone (NFZ) might have been imposed for political reasons that may eventually be removed by diplomacy. The Sprocket's old NFZ FoM can now be removed at the command center. The two sprockets used in this research had six and eight FoMs, respectively.

This research provided useful insights into human decision making in complex systems. It examined multiple user interfaces: tabular text-based, graphical bar charts, analog dials and visual thinking sprockets. The text tables were similar to the standard user interface used in the Predator's UAV ground control station VIT. These tables are long, tedious and required a hardcopy, with the potential of easily misinterpreting the

data. The bar charts and analog dials are a readily available graphics format that could very easily be implemented in an attempt to create a "visual" interface. Finally, the visual thinking widget was developed to present the data to the operator in a format easily interpreted by the operator.

This research also introduced a method to extract multiple levels of "correctness" for a rank order selection. Traditional research presents the results of the rank ordering as either correct or incorrect. This research presented a simple method to expand the incorrect ranking into multiple levels of wrongness. Since a three-way rank order has six permutations and only one absolutely correct, we examined a method to expand the five remaining permutations into three levels of "wrongness" or "less correct".

Besides the "Correctness" of answers, the Response Time needed to be examined. This led to a discussion of when to choose the better display. If two displays are being considered (A and B) and both of them are new displays, then the implementation of one over the other is straight forward. There are nine possibilities shown in Table 7.1. Tie breakers can be personal preference, program requirements, coin flip, etc.

Table 7.1 Simplified Decision Matrix when choosing between two displays, - means same

Display A		Display B		Result	Decision
Correctness	Response Time	Correctness	Response Time		
Better	Better			Strong A	**Display A**
		Better	Better	Strong B	**Display B**
Better			Better	Tie	**Weighted Decision**
	Better	Better		Tie	**Weighted Decision**
Better	-		-	Prefer A	**Display A**
-	Better	-		Prefer A	**Display A**
-		-	Better	Prefer B	**Display B**
-		Better	-	Prefer B	**Display B**
-	-	-	-	Tie	**Weighted Decision**

However, if one of the displays is the current existing display, then tradeoffs have to be considered; hence the weighted decision in Table 7.1. The expense of replacing an existing display with a new display must balance cost in acquisition, deployment and training. Furthermore, the cost of overcoming the human nature to avoid change when a current solution "works" needs to be addressed in an objective manner.

In the case of this research, no standard display exists. If one display is significantly faster but not statistically different in "Correctness", it still might be prudent to change to the faster display. Faster display is the mean response time for each display type. The designer of the decision support system may weigh a bracketing of the speed more important than the mean response time.

Mathematically, the expanded decision rules needed to objectively determine which Display Type to choose are given by

$$Decision = \max \sum_{i=0}^{n} \left(w_1 Correct_i + w_2 \mu_{ResponseTime_i} + w_3 \sigma_{ResponseTime_i} \right)$$

Where μ and σ are the statistical mean and standard deviation, and w is their associated weights. The final decision is a linearly weighted sum of the response times.

This discussion was prompted by the results of the second experiment, in which the Sprocket design was the best or tied for best display. Since this research was supported by a for profit company, recommendations needed to be presented to the customer. Hence the decision table in Table 7.1. Using this decision table, the clear preference is the Sprocket Display.

Although the Visual Thinking Sprocket shows promise, further research is necessary to examine its limitations. The Visual Thinking Sprocket was implemented as a web applet, and it was animated by the WeatherBug.com data feed, thus showing a moderate amount of generic appeal. However, this was by no means a complex, time sensitive domain. The design of the bar chart display was limited by the standard Java® graphics library. Designing the bar chart display from scratch may have affected the results with respect to accuracy and speed. However, the Java® library is optimized, while our display was not and the Visual Thinking Sprocket was still faster.

Furthermore, screen metrics comparing the absolute screen space needed by the various displays should be performed. While the Visual Thinking Sprocket can be resized dynamically, the "readability" of the displays may limit the minimum size. The Analog Gauges were made as large as possible to allow eight FoMs, but still had readability issues. This interaction of readability and size must be researched further.

Appendix A – Experiment 1: Data Displayed in Three Templates

The following pages illustrate the three template versions displayed. The first page was used in the Training Slide to give the subjects a sample of the same data displayed in the three templates – Bar Chart, Tabular and Visual Thinking Sprocket.

The top figure illustrates the bar chart template of the data. If we look at the first bar chart, the dimension/axis name is called "Survivability" and represents the probability of survival. The label on the right reports directionality of the dimension – either the largest bar is best (illustrated by the ↑ arrow) or the smallest bar is best (illustrated by the ↓ arrow). In this experiment, the first dimension is the only one in which largest is best. The label beneath each chart reports the weighting of each dimension. The Survivability dimension has a weighting factor of 40%, while the Defense Network Tracks only has a weighting of 9%. On each vertical axis is the range of interest for each dimension. For each of the probability dimensions the range is held constant between [0.0 .. 1.0]. Otherwise, the upper limit was automatically generated by the charting program (Microsoft Excel® 2007). The final chart, EW Track Radar Probability of Detection, has the Legend for all of the charts – Blue is Route 1, Green is Route 2, Yellow is Route 3, and Red is the Criterion. Examining the Survivability bar chart tells the operator that bigger bars are better, all three routes surpass the criteria and Route 1 has the best probability of survival.

The second template illustrating the layout of the data is the Tabular form. The top row states the columnar headings: Weighting (of each dimension), the Figure of Merit (dimension), Route 1, Route 2, Route 3, Criterion (for each dimension), and directionality (whether higher numbers are better [+] or lower numbers are better [-]). Any values within the route columns that are black have exceeded the criterion, while those in red did not exceed the criterion.

Finally, the third template makes the same information available in the Visual Thinking Sprocket. Each pie slice is labeled with its dimension name for each route. Each dimension is normalized to be the same diameter. The directionality is fixed so that the best possible value always is on the edge of the sprocket, *i.e.*, large is always best. The weighting of each dimension is illustrated by the arc of the pie slice. The criterions for all of the dimensions are normalized to the same radius and are represented by the red circle. Any dimension that exceeds its criterion is colored blue, while those that fail are pink/red.

Appendix B – comments for experiment 1

The following comments are correctly transcribed from the subjects questionnaires on experiment one. [ ] indicate difficult terms to interpret or read in the handwriting.

Any comments about the Text Table block of the experiment?

- Very easy to distinguish between close choices. The red was also very helpful.
- Since certain criteria were weighted it differently, four missions with similar numbers or number of failures it became difficult sometimes to distinguish which option was best. I almost wanted to multiply the row data by their weights and some the results to get a better feeling of what was REALLY best.
- This was a little better than the bar graphs and those deficiencies were coded in red, but it was still hard to compare the routes while taking into account the different weights. No, the tables are not comfortable. They gave me [a] headache.
- A lot of comparing each and looking but all on tables… Better than the bar one. Need to look at them, but not too bad.
- It would help to invert probability of survival to match the others (lower is better).
- The table lacked a sense of scale. His hard to tell how bad things are except in reference to the criterion.
- When all of the three routes are possible it takes time looking at the table and deciding which is the best route.

Any comments about the Bar Chart block of the experiment?

- Making the Headings match the question would make it easier to read / understand. Charts were very confusing. Color coding was helpful, but somewhat overwhelming.
- If survivability were replaced by "probability of loss" so that it went the same way you would have been helpful. Also, I think of the routes in reference to their color rather than number. Waiting is difficult in this format.
- This was very confusing and time consuming.
- It's hard to come up with a decision. It's slightly confusing.
- Tedious compare to the other two.

- It was easy to understand, but all those best choices per chart seemed to blend together, so that some right choices are very hard to choose between. The color was very helpful in choosing between the routes.
- Deftly the most difficult to decipher as far as total merit. It might be the easiest to decipher for a single value, however. To determine the "best" mission route I was left and basically deciphering of each category was a pass / fail and counting the number of failures. Probably not the best approach, but I couldn't think of any easier / better way.
- A lot more difficult than the sprocket. It was hard to remember that on survivability hire is better but lower as better on the other two. He was also difficult to factor in weights since the bars did not change and thickness or area to reflect a change in wait.
- When lower values were better, the difference of the values with the largest values should be displayed instead of the values themselves!
- There's a lot to think about when reading it. I have to look at each one again and again to answer the question. Need to recognize if higher is better or which one is which. There are also 7-8 bar graphs to look at.
- Weighting factor was not represented properly.
- It was difficult to read the data off the bar chart. You had to keep looking / staring at the charts and thinking of different alternatives. It was quite tedious, also when compared to the other two (Sprocket & Text Table).

   Any comments about the Sprocket block of the experiment?

- And bigger size of sprocket and probably use separate color instead of think has a son something carry as difficult a notice (as in values which were very small).
- This was the best thing to make quick decisions. It was easy to see the most important and easy to understand the criteria. I think I did the best on this. I think it is a good idea to use the same (highest / lowest is better) because it takes extra time to think cool weather that one should be higher or lower. It's not hard, but it's an extra hoop to jump through.
- Difficult to notice small differences quickly but easy for large.
- Easy to quickly determine the relative values of each of the parameters and the all overall weighted area.
- There are only three circles instead of so many bars. It was easier to read, see, and understand. Taking something out of which one is better – this allows me to just take a look and not have to rethink the graphs.
- It would be good if read was used in place of thinking. Red is always associated with danger and it's kind of always known for danger everywhere.

- The ascending or descending order of criteria is important like probability of survival, etc. [min/max] but for a [non pilot] plane doesn't matter, but is confusing at first to make decision as an engineer (mechanical) safety came into mind first.
- It is easier to make a quick decision when all mission options are coated the same (more is better). Text for `samTracking` overlaps the pink area so was hard to distinguish. Although this was the easiest, it helps to have numbers available when they are close in value.
- It's very clear and helpful.
- I believe it is easy to identify the route and is not confusing. Gives the user quick decision-making ability.

Any comments about the overall experiment?

- Very interesting experiment. The sprocket reduces the decision-making time by a large extent compared to the [Text] Table and Bar Chart.
- The table decision data was the best according to me and was pretty fast and easy to understand. The color coding of the bar chart was a little bit confusing though.
- Good experiment. I would've thought the bar chart would be best.
- I really like the sprocket idea as it gives a clear and easy visual cue by using volume to quantify "strength".
- Very good test.
- I like this experiment. This pretty good. I would like to participate, if such set of experiments are held in [the] future.
- It's hard for me to read the bar chart. Because the number [were crossed out???].
- The sprocket is the way to go! Bar graph is not useful at all. I believe the bar graph is actually a step backward from the data table.
- I had bargained tables first and then finally [the] sprocket. It is sometimes difficult to come out of state of mind that bigger is always better.
- The bar chart was the difficult one and it took a lot of time. [I] felt a little strain doing that. The sprocket was more handy with everything pictured. In bar chart, you had to come back again to the different charts when comparing routes.
- Visual learning has always been my preferred method of learning and this kind of proves that.
- Went smoothly, but you need to get a new projector without the lack of color option.
- Nice work.
- It [is] a good one and the sprocket comes out as a clear winner.

References

Army (2000). Air Cavalry Squadron and Troop Operations. <u>Field Manual 1-114</u>. Army. Washington, DC, Department of the Army**:** 5-9 thru 5-21.

Arnheim, Rudolf (1969). <u>Visual Thinking</u>, University of California Press.

Barnett, B. J. and Christopher D. Wickens (1988). "Display proximity in multicue information integration: The benefit of boxes." <u>Human Factors</u> **30**(1): 15-24.

Besco, Robert O., Dan Maurino, Martin H. Potter, Barry Strauch, Richard B. Stone and Earl Wiener (1994). <u>Unrecognized Training Needs for Airline Pilots</u> Human Factors and Ergonomics Society Annual Meeting Proceedings: Aerospace Systems, Human Factors and Ergonomics Society.

Bonabeau, Eric and Guy Théraulaz (2000). Swarm Smarts. <u>Scientific American</u>**:** 72-79.

Boyd, Colonel John. (1996). "Boyd's OODA Loop."   Retrieved 18 July, 2006, from <u>http://www.d-n-i.net/boyd/boyds_ooda_loop.ppt#287,1,The</u> OODA "Loop" Sketch.

Casson, R W (1983). "Schemata in Cognitive Anthropology." <u>Annual Review of Anthropology</u> **12**(1): 429-462.

Cleveland, William and Robert McGill (1984). "Graphical perception: Theory, experimentation and application to the development of graphical methods." <u>American Statistical Association</u> **79**(387): 531-554.

Cummings, Missy L. (2004). <u>Human Supervisory Control of Swarming Networks</u>. 2nd Annual Swarming: Autonomous Intelligent Networked Systems Conference.

Cummings, Missy L. and Stephanie Guerlain (2004). An Interactive Decision Support Tool for Real-time In-flight Replanning of Autonomous Vehicles. <u>AIAA 3rd "Unmanned Unlimited" Technical Conference</u>. Chicago, Illinois, American Institute of Aeronautics and Astronautics**:** 1-9.

Cummings, Missy L. and D. Morales (2005). "UAVs as Tactical Wingmen." <u>Unmanned Systems</u> **23**(1): 5.

Dekker, S. W. A. and D. D. Woods (2002). "MABA-MABA or abracadabra? Progress on human-automation coordination." <u>Cognition Technology, and Work</u> **4**: 240-244.

Donovan, Sarah-Louise and Tom Triggs (2006). Investigating the Effects of Display Design on Unmanned Underwater Vehicle Pilot Performance. Defence Science and Technology Organisation, Maritime Platforms Division.

Dorigo, Marco, Eric Bonabeau and Guy Theraulaz (2000). "Ant algorithms and stigmergy." <u>Future Generation Computer Systems</u> **16**(8): 851-871.

Durso, F., M. K. Bleckley and A. Dattel (2007). "Does situation awareness add to the validity of cognitive tests?" <u>Human Factors</u> **48**: 721-733.

Edlund, C. and M. Lewis (1995). <u>Fidelity of dynamics in integrated displays</u>. Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on.

Endsley, M. R. and E. O. Kiris (1995). "The Out-of-the-Loop Performance Problem and Level of Control in Automation." Human Factors **37**(2): 381-394.

Endsley, Mica R. (1988). Design and evaluation for situation awareness enhancement. Human Factors Society 32nd Annual Meeting Santa Monica, CA, Human Factors Society.

Endsley, Mica R. (2006). Expertise and Situation Awareness. The Cambridge Handbook of Expertise and Expert Performance. K. A. Ericsson, N. Charness, P. J. Feltovich and R. R. Hoffman. New York, Cambridge University Press**:** 633-652.

Felder, R.M. and L.K. Silverman (1988). "Learning and Teaching Styles in Engineering Education." Engr. Education **78**(7): 674-681.

Ferber, Jacques (1999). Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence, Addison-Wesley.

Ferreira, Jennifer, Pippin Barr and James Noble (2005). The Semiotics of User Interface Redesign. Sixth Australasian Conference on User Interface.

Franke, Jerry L., Vera Zaychik, Thomas M. Spura and Erin E. Alves (2005). Inverting the Operator/Vehicle Ratio: Approaches to Next Generation UAV Command and Control. Unmanned Systems North America 2005 Conference. Baltimore, MD, Association for Unmanned Vehicle Systems International and Flight International.

Gaudiano, Paolo, Benjamin Shargel, Eric Bonabeau and Bruce T. Clough (2003). Control of UAV Swarms: What the Bugs Can Teach Us. AIAA "Unmanned Unlimited" Systems, Technologies, and Operations—Aerospace, Land, and Sea, San Diego.

Giffen, W. and T. H. Rockwell (1987). A Methodology for Research on VFR flight into IMC. Fourth International Symposium on Aviation Psychology. Columbus: The Ohio State University, R.S. Jensen.

Gunter, SSG Christopher and SGT Lytle (2005). "The Operator's Perspective" an informal Q&A session with UAV operators. n. A. H. F. o. UAV's. Mesa, AZ, CERI.

Hart, Sandra G. (2006). NASA-TASK LOAD INDEX (NASA-TLX); 20 YEARS LATER. Human Factors Engineering Society, Florida, HFES.

Hildebrandt, Michael and Michael Harrison (2002). The Temporal Dimension of Dynamic Function Allocation. 11th European Conference on Cognitive Ergonomics (ECCE 11).

Hildebrandt, Michael and Michael Harrison (2003). Putting Time (Back) into Dynamic Function Allocation. Human Factors and Ergonomics Society (47th), Denver, Colorado.

Holland, Owen, John Woods, Renzo De Nardi and Adrian Clark (2005). Beyond Swarm Intelligence: The UltraSwarm. IEEE Swarm Intelligence Symposium SIS2005, Pasadena, CA, IEEE.

Hollnagel, Erik and David D. Woods (1999). "Cognitive Systems Engineering: New wine in new bottles." International Journal of Human-Computer Studies **51**(2): 16.

Jenkins, Joseph Christopher (2007). The Effect of Configural Displays on Pilot Situation Awareness in Helmet-Mounted Displays. PhD Dissertation, Wright State University.

Johnson Controls Inc. (2001). Starfield Display User's Guide, Johnson Controls, Inc.

Karim, S., C. Heinze and S. Dunn (2004). <u>Agent-based mission management for a UAV</u>.

Klein, Gary (1993). <u>Naturalistic Decision Making: Implications for Design</u>. Wright Patterson AFB, CSERIAC.

Klein, Gary (2009). Conversation about Situation Awareness weaknesses. M. J. Findler. Dayton.

Lewis, Michael, Jumpol Polvichai, Katia P. Sycara and Paul Scerri (2005). Scaling-up Human Control for Large UAV Teams. <u>The Human Factors of Remotely Piloted Vehicles</u>. N. Cooke. New York, Elsevier.

Lewis, Michael, Jijun Wang and Paul Scerri (2006). Teamwork Coordination for Realistically Complex Multi Robot Systems. <u>NATO Symposium on Human Factors of Uninhabited Military Vehicles as Force Multipliers</u>.

McCormick, E. J. (1976). <u>Human Factors in Engineering and Design</u>. New York, McGraw Hill.

McKim, Robert H. (1972). <u>Experiences in Visual Thinking</u> Belmont, California, Brooks/Cole Publishing Company.

McNeese, Michael, Holly Bautsch and S. Narayanan (1999). "Tradeoffs in cognitive field research: Case study of modeling fighter pilot performance." <u>International Journal of Cognitive Ergonomics</u>.

Miller, R. A., Ed. (1985). <u>A system approach to modeling discrete control performance</u>. Advances in Man-Machines System Research. Greenwich, CT, JAI Press.

Mitchell, Christine M. and R. A. Miller (1986). "A discrete control model of operator function: A methodology for display design." <u>IEEE Transactions on Systems, Man, and Cybernetics</u> **16**(3): 353-369.

Nadin, Mihai (1988). "Interface design: A semiotic paradigm." <u>Semiotica</u> **69**(3/4): 33.

Narayanan, S., Narasimha Rao Edala, Jonathan Geist, Patchigolla Kiran Kumar, Heath A. Ruff, Mark Draper and Mike Haas (1999). "UMAST: A Web-Based Architecture for Modeling Future Uninhabited Aerial Vehicles." <u>Simulation</u> **73**(1): 29-39.

Narayanan, S., Heath A. Ruff, N.R. Edala, J.A. Geist, K. Patchigolla, M. Draper and Michael Haas (2000). "Human-integrated supervisory control of uninhabited combat aerial vehicles." <u>Journal of Robotics and Mechatronics. Special Issue on "Intelligent Control in Coming New Generation</u> **12**(6): 1-12.

Parasuraman, Raja, Thomas B. Sheridan and Christopher D. Wickens (2000). "A Model for Types and Levels of Human Interaction with Automation." <u>IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans</u> **30**(3): 286-297.

Pomerantz, J. R. (1986). Visual form perception: An overview. <u>Pattern recognition by humans and machines</u>. H. C. Nusbaum and E. C. Schwab. Orlando, FL, Academic Press. **2:** 1-30.

Pu, Pearl and Denis Lalanne (2002). <u>Design visual thinking tools for mixed initiative systems</u>. Proceedings of the 7th international conference on Intelligent user interfaces, San Francisco, California, USA ACM.

Rothrock, Ling, Kimberly Barron, Timothy W. Simpson, Mary Frecker and Chris Ligetti (2006). "Applying the Proximity Compatibility and the Control-Display Compatibility Principles to Engineering Design Interfaces." Human Factors and Ergonomics in Manufacturing **16**(1): 61-81.

Rouse, William B., Norman D. Geddes and Renwick E. Curry (1987-1988). "An Architecture for Intelligent Interfaces: Outline of an Approach to Supporting Operators of Complex Systems." Human-Computer Interaction **3**: 87-122.

Ruff, Heath A., Gloria L. Calhoun, Mark H. Draper, John V. Fontejon and Brian J. Guilfoos (2004). Exploring Automation Issues in Supervisory Control of Multiple UAVs. Human Performance, Situation Awareness, and Automation Technology, Daytona Beach, FL.

Ruff, Heath Anthony (2000). The Effect of Automation Level and Fidelity on the Human Supervisory Control of Multiple Remotely Operated Vehicles. Masters of Engineering, Wright State University.

Scerri, Paul, Yang Xu, Elizabeth Liao, JJustin Lai, Mike Lewis and Katia Sycara (2004). Coordinating very large groups of wide area search munitions. Recent Developments in Cooperative Control and Optimization. Dordrecht, NL, Kluwer Academic Publishers.

Sharek, David. (2010). "http://www.playgraph.com/." Retrieved October 1, 2010, 2010.

Sheridan, T. B. (1997). Supervisory Control. Human Factors and Ergonomics Handbook. G. Salvendy. New York, Wiley**:** 1295-1327.

Silverman, Fran (2006). "Learning styles." District Administration **42**(9): 70-71.

Singers, Robert R. and Linda S. Endres (1996). Starfield: The Next Generation. 3rd Annual Symposium on Human Interaction with Complex Systems. Dayton, OH, IEEE.

Spence, Robert (2001). Information Visualization. Essex, England, Pearson Education Limited.

SRA International Inc. (2005, 18 April). "UAV Forum." Retrieved 19 April, 2005, from http://www.UAVForum.com.

Tambe, Milind (1998). "Implementing Agent Teams in Dynamic Multi-agent Environments." Applied Artificial Intelligence **12**.

Tenney, Yvette J., MariIyn Jager Adams, Richard W. Pew, A. W. F. Huggins and William H. Rogers (1992). A Principled Approach to the Measurement of Situation Awareness in Commercial Aviation, NASA**:** 38.

Tufte, Edward R. (1997). Visual Explanations. Cheshire, Connecticut, Graphics Press.

Tufte, Edward R. (2006). The Visual Display of Quantitative Information. Cheshire, Connecticut, Cheshire Publishing.

Valente, Luis, Clarisse Sieckenius de Souza and Bruno Feijó (2008). An exploratory study on non-visual mobile phone interfaces for games. Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems. Porto Alegre, RS, Brazil, Sociedade Brasileira de Computa\&\#231;\&\#227;o**:** 31-39.

Ware, Colin (2004). Information Visualization: Perception for Design. San Francisco, CA, Elsevier.

Wickens, Christopher D. and C.M. Carswell (1995). "The proximity compatibility principle: Its psychological foundation and relevance to display design." <u>Human Factors</u> **37**(3): 473-494.

Wikipedia. (2007). "MQ-1 Predator." Retrieved 4 December 2007, 2007, from http://en.wikipedia.org/wiki/RQ-1_Predator.

Wikipedia contributors. (2010, 12 February 2010 19:27 UTC). "Semiotics." Retrieved 24 February 2010 18:44 UTC, 2010, from http://en.wikipedia.org/w/index.php?title=Semiotics&oldid=343593085.