The Impact of Computer Decision Support on Military Team Decision Making

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Adam Donavon Larson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Caroline Hayes

August 2010

DISCLAIMER CLAUSE: The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government

Acknowledgements

I would like to express my everlasting thanks to my wife Melissa and children, Isabella and Ethan, for their endless prayers and support. I will love you always.

A heartfelt thank you is given to my advisor, Prof Caroline Hayes, for her guidance and encouragement over the course of my time at the University of Minnesota. Your expertise and support will carry on wherever life leads me. Thank you Ma'am!

Appreciation is also offered to Dr. Thomas Stoffregen, Dr. Tom Smith, and Dr. Robin Schaller for serving on my examination committee.

Dedication

This dissertation is dedicated to my Dad, Donavon Larson, a lifelong teacher and coach, who taught me so many important lessons in life.

"One day at a time, with everything you've got!"

Abstract

This dissertation work highlights extremely valuable results regarding significant costs and benefits of using a computer decision aid by analyzing the impact of such a decision support tool on military team decision making. Decision support systems (DSS) are becoming increasingly popular as an approach to aid decision makers in making better decisions in a more efficient and effective manner. However, DSSs have both costs and benefits in their utilization, and there is no guarantee that a DSS will actually improve decision making or problem solving performance. This work shows that although a DSS has many advantages and can facilitate user problem solving, brittle DSS behavior can significantly degrade user decision making.

The primary goals of this work are to improve scientific understanding of situations in which DSSs may improve decision making performance and those where the use of a DSS may actually degrade performance. Specifically, the heart of this work focuses on understanding and measuring the performance benefits and costs of a solution generating DSS on individuals versus teams, and on situations in which the DSS produces "brittle," or questionable solutions. Understanding the impact of brittle behavior is especially important given the domains in which DSSs are often utilized, including military, medical, and business operations. The results of decisions in these areas greatly impact dollars and most importantly, human lives, that may be saved or lost.

The decisions teams make in military situations play a vital role in determining the success or failure of operations. Decision support in this study was provided by a component of a DSS tool called Weasel. A previous study in 2004 analyzed Weasel with respect to *individual* decision makers' performance and behavior [9]. This study

analyzed *team* behavior and performance in a military context with military personnel working together in three person teams. The primary questions addressed by this work are: What is Weasel's overall impact on team versus individual performance and what is the effect on user performance when Weasel exhibits brittle behavior? Brittle behavior refers to the automated decision tool offering questionable, low quality courses of action for a given situation. As all DSSs will at sometime or another exhibit some degree of brittle behavior, the impact of such behavior on user decision making is vitally important.

The results showed brittle behavior does indeed negatively impact user decision making behavior, and that individuals and teams demonstrated the same levels of performance with the use of the automated decision tool. The results of this experiment will help researchers and military personnel to better understand when it is appropriate to use decision support and to better understand both the benefits and the costs in team decision making by assessing when the DSS tool facilitated improved decision making and when performance was hindered by the tool. Additionally, information may be gained regarding situations where computer support and automation use may degrade performance.

Table of Contents

Page

Page

List of Figures

Page

Chapter 1 - Introduction

The Impact of Computer Decision Support on Military Team Decision Making

This dissertation describes an experiment designed to analyze the use of a solution-generating decision support system (DSS) tool, Weasel, and its effects on team and individual decision making performance as well as how brittle DSS behavior impacts decision makers. There is a wide array of benefits and costs when using a DSS. Benefits can include improved decision making, time saved, highly complex problems being analyzed faster and to a deeper level of understanding, and new solutions being proposed which the human decision maker may not have thought of. However, there are also significant costs of such a system, including adopting a poor solution generated by the DSS which can lead to lives or money being lost, and over-reliance on the automated tool leading to human decision makers not critically or appropriately assessing a problem.

DSSs are becoming increasingly popular in various domains as a way to aid decision makers in making better decisions in a more efficient and effective manner. Such systems are accepted as decision-making aids in a wide variety of high-criticality decision making tasks, many with life and death consequences. From economics to medical problems and aircraft navigation to military operations, the implementation of DSSs is growing. Thus, it is important to understand the effect a DSS may have on human decision makers.

However, there is no guarantee that a DSS will improve decision making or problem solving performance. In fact, is it a widely held belief in this area of study that all DSSs will exhibit brittle behavior at one time or another [9, 20]. Brittleness refers to the tool offering low quality, questionable courses of action for the decision maker to consider. If this is indeed true, it is important to understand how such brittleness impacts the human decision maker and the ultimate decisions that are being made.

There are many challenges inherent in complex decision making domains such as military operations, economics, business, and medical operations. Decision makers in many of these areas are presented with high stress levels, potential information overload or lack of information, various sizes and social dynamics of teams or groups involved in the decision making process, dynamic operational environments, and vital objectives that must be met. These factors can drastically impact the quality and types of decisions made and hence, there is a need to research such environments to determine if, and how, decision support systems may be useful to aid decision makers. DSS tools may ease the decision maker's burden in complex operational environments.

Specifically pertaining to the military domain, the advanced technological tools used by today's military forces constitute a need for further understanding the complex interaction between the human operator and the automated system being employed. Automated systems often enhance what can be achieved in military tactics, up to the limitations of that system as well as the human(s) operating or monitoring the system. Previous work has been conducted analyzing military decision making, the impact of military DSS tools, and team dynamics. However, the major contribution of this study is that it combines analysis of those three factors by offering greater understanding with regards to military team decision making when using a DSS.

This study's primary goal is to improve scientific understanding of situations in which DSSs may improve decision making performance and those where the use of a DSS may actually degrade performance. Specifically, the focus of this work is on understanding and measuring the performance benefits and costs of a solution generating DSS on individuals versus teams, and on situations in which the DSS produces "brittle," or questionable solutions. The implementation of DSS technology can be significantly enhanced if we have greater knowledge regarding when and how the system improves or degrades performance, and how brittle DSS behavior impacts decision making.

Weasel is a DSS designed to assist military planning staff in generating courses of action (COAs) for ground forces. Weasel is the tool utilized in this study since it is present and available for use at the University of Minnesota. However, as the system is adequately representative of DSS tools as a whole, the results and lessons learned from this study can be generalized to team decision making and DSS applications as a whole. There remains to be somewhat limited knowledge regarding the use of intelligent automated assistant tools in team decision making and military operations, and no such study has been conducted specifically analyzing team decision making with the use of Weasel.

This work is motivated by various issues that if analyzed and understood further, may enhance the effectiveness of DSS use and improve decision making in complex domains (including military operations). One issue worthy of analysis is that of individual versus team decision making. Decisions across domains are sometimes made by one decision maker and in many other situations a team of personnel working together decides on a proper course of action. These different decision making structures take place in the military as well as in civilian sectors such as finance and healthcare. Analyzing how a DSS may impact individual and team decision making can significantly impact how the system may be implemented and utilized to maximize decision making effectiveness. A key result from the author's research in 2004 found individual performance did improve when using the DSS [9]; it is worth analyzing if the same holds true for teams working with the DSS.

Another significant issue needing analysis is regarding the impact on user behavior and performance when a DSS demonstrates "brittleness". Brittleness here refers to the DSS generating courses of action that are questionable or low-quality for the given scenario parameters. For example, suppose an aircraft has two possible flight paths it may follow to arrive at a desired location. One path has a clear weather pattern while the other path goes through a weather system which would cause turbulence and potential flight delays. If the DSS would recommend the aircraft take the flight path through the poor weather system, this DSS behavior would be considered brittle. Previous research has shown brittleness by a DSS may degrade user performance [20]. Brittleness is a key aspect of DSS behavior to analyze given the serious implications of following such poor courses of action. In the military or healthcare domains, lives may be lost. In a finance DSS application, millions of dollars may be at stake depending on the decision made. The degree on which users rely on the automated tool for potential solutions and the influence of DSS brittleness on user behavior and performance must be understood.

Similarly, the impact of how a problem is solved or when COAs are considered may offer useful insight into how to best utilize the DSS in problem solving endeavors. Previous work by Smith, McCoy and Layton analyzed the impact of presentation order on problem solving behavior [20]. Presentation order refers to when the DSS COAs are presented to the user in the problem solving or decision-making process. The work by Smith found users of a flight planning test-bed DSS had a tendency to choose the computer-generated solutions more when they are shown the computer COAs first, before generating any COAs of their own [20]. Various methods of presentation order may facilitate reducing the decision bias users may inherently have when working with a DSS. In military operations it would be very important to know the propensity users may have to follow the DSS recommendations, especially when the DSS may demonstrate brittle behavior as discussed previously.

This research addressed these challenges by testing military team decision making in problem solving scenarios. The study also gathered data regarding user attitudes towards trust in automation and personal capabilities. By analyzing problem solving performance and corresponding questionnaire data, the research attempts to improve understanding of subject behavior and Weasel's impact on that behavior. Based on the motivations discussed previously, the following specific research questions were addressed:

1. a. When Weasel exhibits brittle behavior, are user's decisions impacted by that brittleness (i.e. do users choose the flawed/brittle solutions generated by the DSS)?

   b. Does ECOA quality decline when Weasel exhibits brittle behavior?

2. a. Does Weasel help *teams* to produce better quality COAs?

   b. Does Weasel help users working in teams more than individuals?

3. a. Does presentation order impact preference toward computer solutions?

   b. Does presentation order impact performance?

4. How do user attitudes (with regards to satisfaction, confidence and conflict management) relate to performance or decision behavior?

Answering these questions will provide insight into the usefulness of Weasel and increase understanding of DSS user behavior and performance in a team military context. The significance of this research may be seen in greater understanding of brittle DSS behavior, the impact of that brittleness in important decision making domains, improved military training and operations as well as researchers (i.e. those interested in decision making, DSS, and automation) and military personnel (i.e. commanders, strategists, plans officers) gaining greater overall knowledge in the areas of decision making and use of automation.

Chapter 2 - Domain Background

Decision support systems are computer tools designed to support the user's decision making processes and ultimately, to aid the user in making the best possible decision. DSS research typically focuses on how the system "can improve the efficiency with which the user makes a decision and improve the effectiveness of that decision" [19]. Although DSSs were initially designed for individual use, group decision support systems (GDSS) evolved around 1985 to facilitate team problem solving and shortly thereafter, executive information systems (EIS) broadened the application of DSS to the corporate level [19]. Prior research has shown a DSS can be a tremendously useful tool in domains where various complex decisions are made (areas such as medical, business management, and economics to name a few) and they typically contribute to better decisions in both individual and group applications [9, 5, 20, 28]. They are innovative in many aspects as they are often designed with artificial intelligence capability (typically using thousands of complex algorithms), they can process large quantities of data in very short time periods, and their design often promotes different levels of interaction which is determined by the user. The human-computer interface can be thought of as allowing various degrees of interaction as described by Sheridan and Verplank, where level 1 represents the "whole task done by human except for actual machine operation" (described as a manual task) up to level 10, where the "computer does everything autonomously," which is described as an automatic task [18]. Depending on the DSS design and level of human-computer interaction, the various levels of automation can facilitate task completion and decision making.

A particular DSS developed and analyzed at the University of Minnesota in relation to military decision making is called Weasel. Weasel is a DSS designed to assist military planning staff in generating courses of action (COAs) for ground forces. Weasel is one component of a DSS "Intel Tool Kit" developed in 2003 at the University of Minnesota. It is programmed in C++ and uses a constraint search methodology to generate and evaluate plans of enemy and friendly military forces [17]. The Intel Tool Kit consists of Weasel, the enemy course of action (ECOA) generator, Fox, which is the friendly force course of action (FCOA) generator, and CoRaven, an intelligence analysis tool [14]. Another key component of the tool kit is the map function which displays a topographic type area map on which displays can be shown.

The map display in Figure 1 provides a visual depiction of potential enemy actions. Also integrated into the map display are various intelligence components, including avenues of approach (AAs) and lines of defensible terrain (LDTs).

An AA is represented by a broad arrow situated on the horizontal axis, pointing to the right. In Figure 1 there are two AAs displayed, axis white is to the north of axis red. An AA is a route on which military units can move in order to attack or defend. The direction of the arrow shows the direction of force movement. There can be between 2 and 5 AAs in this simulation as selected in the ECOA screen (right column in Figure 2).

The thin vertical lines on the map represent LDTs. There are 5 LDTs as depicted on the map in Figure 1. The LDTs have been labeled left to right as LDT 1-5. LDTs are potential defense positions used to determine depth of forces and to place forces at the

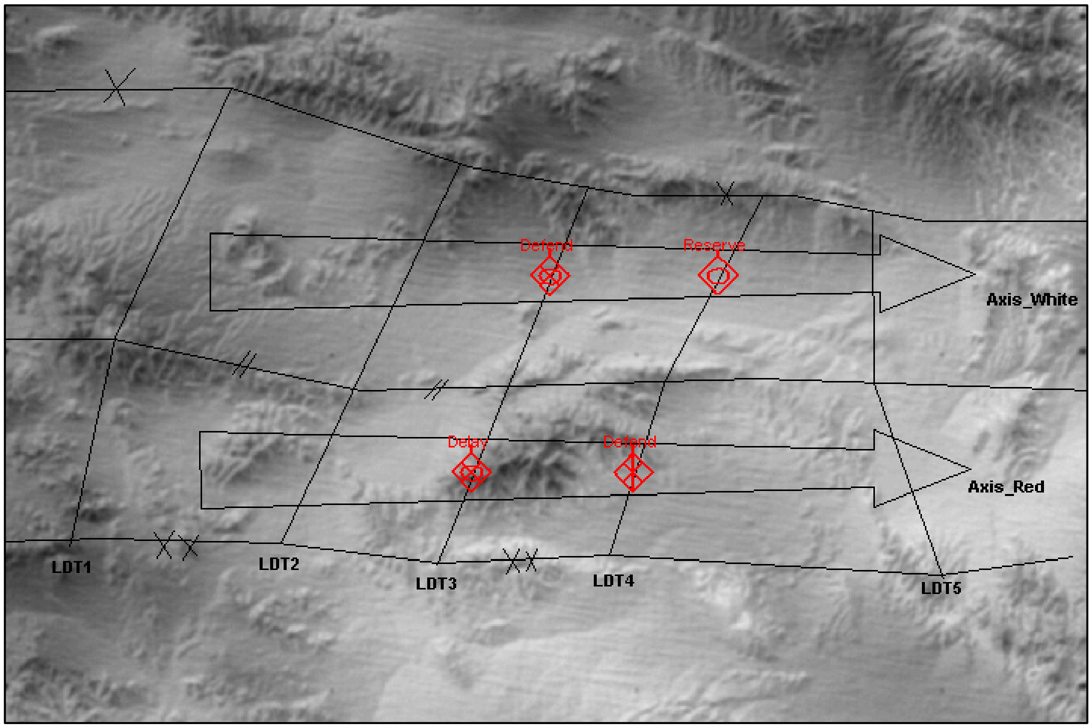


Figure 1. Map display example with enemy forces depicted

intersections of LDTs and AAs. These LDTs were used to mark placement of forces along the AAs.

The ECOA generator shown in Figure 2 is used to determine the makeup of enemy forces and the scenario at hand. The left side of the ECOA page is used to select the type of units that comprise enemy forces. Units can be a battalion, company, or platoon with the type of each unit being mechanized infantry, armor, or motorized forces. On the bottom left of the page, the mission of each force is determined. Attacking forces can either commit to attack (C), follow and support (F&S), or attack in reserve (R). Defense forces can defend (Def), delay in defense (Del), or defend in reserve (R).

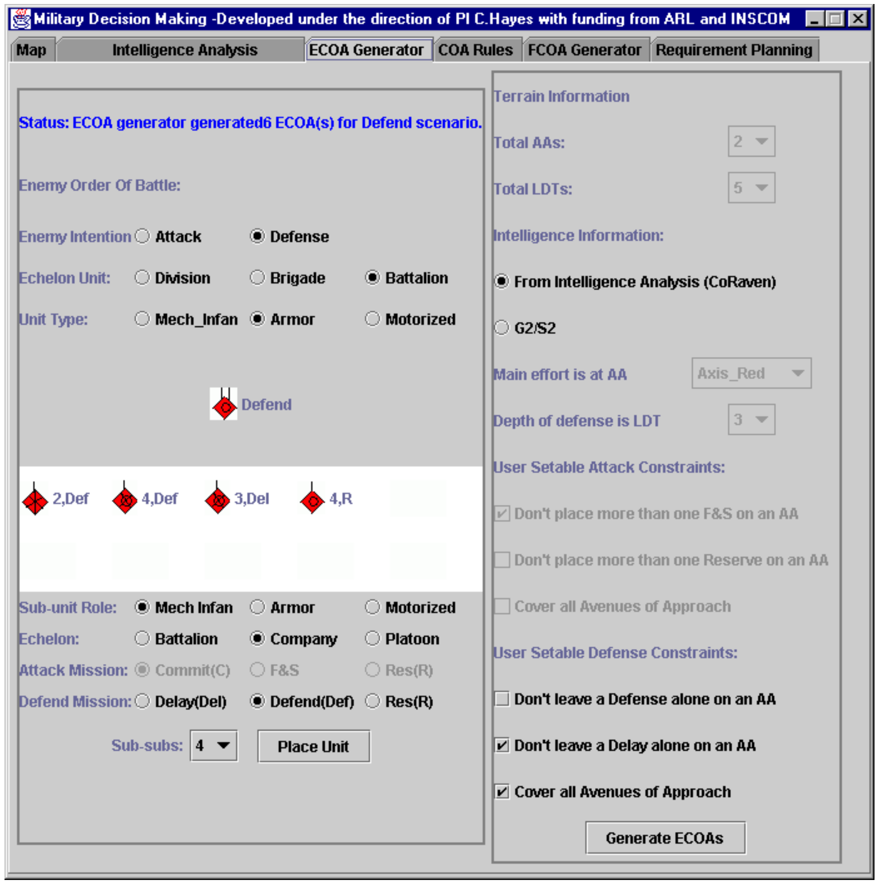


Figure 2. Input screen for Weasel ECOA generator

Once all enemy data or "intelligence" is entered into Weasel, the "generate ECOAs" tab on the lower right portion of the screen is selected. This prompts the DSS to analyze the intelligence and generate potential ECOAs. An example of an ECOA set generated by Weasel can be seen in Figure 3.

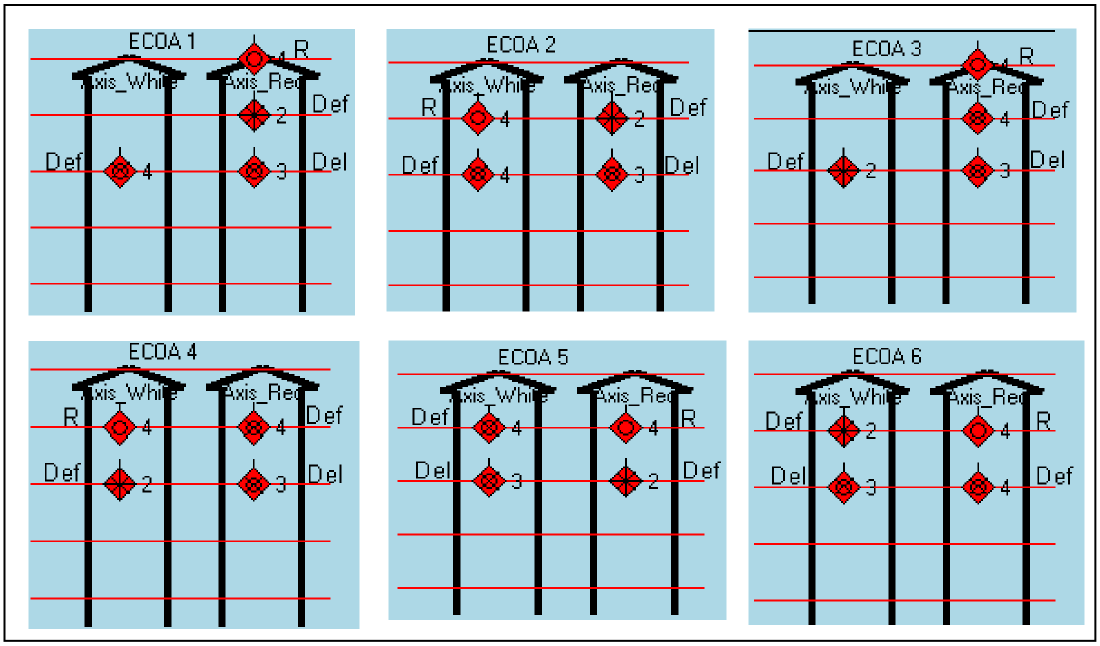


Figure 3. Example of Weasel ECOA set

The U.S. Army symbology (www.fas.org/man/dod-101/army/docs/fm101-5-1/f545-c4a.htm) used in the DSS to depict enemy force information is shown in the following figure [30].

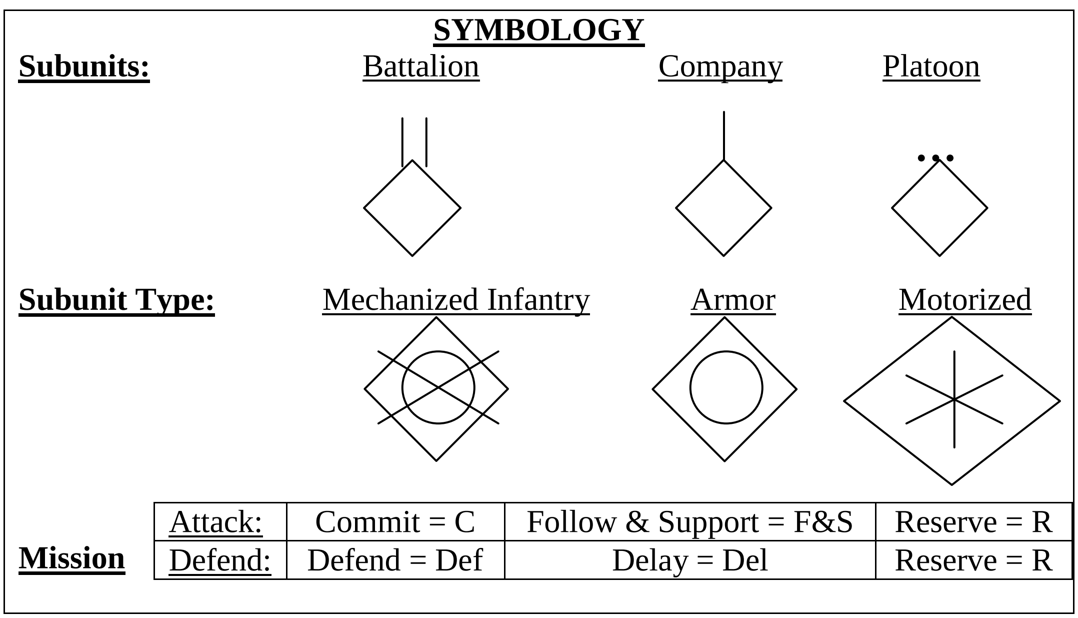


Figure 4. Unit symbology

The 2004 research study [9] analyzed individual military personnel decisions when using Weasel, particularly whether or not subjects chose ECOAs generated by the DSS, themselves, or a combination of the two. A key aspect of the research methodology was that subjects worked alone, generating COAs, analyzing them, and ultimately making decisions by themselves. However, military personnel are most often working together in teams of various sizes. This research aimed to analyze the types and quality of decisions made when groups of people are utilizing a DSS.

Chapter 3 - Literature Review

This literature review consisted of analyzing research pertaining to decision support systems and relevant topics of interest related to such systems. An extensive literature review was conducted analyzing DSS research in relation to group (also referred to as team or collaborative) decision support systems, relationships between decision making and DSS use, and trust in automation. This section consists of a summary of chosen pieces of literature as well as a synthesis of the literature.

3.1 – Literature Summary

Individual Decision Making

- Smith, Philip, C.E. McCoy and C. Layton. "Brittleness in the Design of Cooperative Problem-Solving Systems: The Effects on User Performance" *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 1997. [20]

Smith, McCoy and Layton conducted an experiment exploring flight planning routes of individuals operating commercial aircraft. The study, which utilized a DSS called the Flight Planning Testbed (FPT), found the DSS tool improved users' average problem solving performance in finding fuel-optimal routes for commercial jets, but it occasionally degraded some users' performance when FPT exhibited *brittle* behavior. Brittle behavior is what happens when a DSS produces an inappropriate or inadequate suggestion [20]. In FTP's case, brittle solutions were fuel-optimal, but unnecessarily risky. However, the researchers also found that this effect appeared to be mitigated if subjects did their own exploration of the problem before seeing the computer's solution.

It is with high probability that DSSs, simulations, or mathematical models will exhibit some brittleness because they are *necessarily* simplifications of the real world. Therefore their solutions will produce some degree of error which may or may not be predictable [9]. Thus, given that some degree of brittleness is inevitable, it is important to consider how brittle behavior may impact decision makers in many tasks, and if it can be mitigated.

Also, relating to when DSS information is presented to a user (referred to as presentation order in this study), the FPT research found that presenting an automated recommendation early in a problem solving analysis can significantly impact the decision making process [20]. This impacts the overall situational assessment and evaluation of alternatives. Similar to the work by Smith, McCoy, and Layton, the research described in this study analyzed the decisions users made versus the order in which potential courses of action were presented and developed.

- Larson, Adam D. and Caroline Hayes. "An Assessment of Weasel: A Decision Support System to Assist in Military Planning" *HFES Proceedings*, 2005. [9]

This work describes the study from 2004 analyzing the use of a DSS tool, Weasel, and its effects on military decision maker's performance. Because of the increased complexity and tempo of modern operations, there is great interest in possible use of decision support tools to assist in high-criticality decision making tasks such as military planning. However, because these tasks can impact human safety and may have political ramifications, it is important to clearly understand how they influence human decision making before adopting such tools. Specific goals of this study were to identify whether use of Weasel affects users performance by 1) improving quality of ECOAs, 2) changing novices' performance more than experts', 3) decreasing quality when Weasel produces inappropriate or inadequate solutions and 4) whether showing users computer generated solutions before they have assessed the situation affects their performance.

The results from the study found that:

- Overall there was a significant improvement in quality scores when ECOAs generated *without* Weasel's assistance were compared to ECOAs generated *with* Weasel's assistance ($p = 0.018$). ECOA generated with Weasel were significantly better.
- Weasel improved novices' quality scores significantly ($p = 0.0002$), but did not significantly change experts' quality scores ($p = 0.251$). Furthermore, differences between experts and novices were leveled when both groups used Weasel; there was no significant difference between novice and expert quality scores when using Weasel ($p = 0.366$).
- When the DSS exhibited "brittle" behavior by producing inadequate solutions, *more* subjects' (5 out of 18) tended to repeat the mistake made by Weasel (i.e. omission of ECOAs that "cover" all avenues of approach). Furthermore, three of the five who made the omission were experts. In contrast, only one subject (a novice) made this same mistake when producing solutions manually. This implies that use of Weasel may have "biased" users towards flawed solution sets when it exhibited brittle behavior.

- Presentation order did not significantly impact user decisions (i.e. viewing the computer generated solutions first did not lead to an increased tendency to chose the computer ECOA set).

Group Decision Making

- Gallupe, R. Brent et al. "Computer-based support for group problem-finding: an experimental investigation" *MIS Quarterly*, 1988. [5]

This lab study examined the effects of GDSS technology on group decision quality and individual perceptions when applied in the context of face-to-face meetings. The task involved was a crisis management task and two versions were run in the trials, one with a higher level of difficulty than the other. The design was a 2 x 2 factorial design with some groups receiving support from a GDSS, while other groups did not. Each group consisted of three subjects. The role of the GDSS was to record, store and display alternatives entered by the group members, display preference rankings of those alternatives as entered by the group, and recording votes for the alternatives entered into the system [5].

Decision quality was significantly better in groups that received GDSS support, which is consistent with past research. The GDSS help was especially evident in groups who received the problem-solving task with a high level of difficulty.

One of the very interesting results from this study found that member decision confidence and satisfaction were lower and inter-member conflict was greater in GDSS-supported groups, despite the fact that performance was greater. This result is contrary to that found by Steeb and Johnston [21] and Turoff and Hiltz [26]. The authors felt the lower confidence levels in this study may have been due to "post-decision apprehension" since the GDSS-supported groups typically generated more alternatives, weighed those alternatives in greater detail, and possibly had a more difficult decision to make. These conflicting results offer the opportunity for additional research endeavors to analyze the issue of user confidence and satisfaction when using a DSS.

- Holloman, Charles and Hal Hendrick. "Adequacy of group decisions as a function of the decision-making process" *The Academy of Management Journal*, 1972. [6]

The introduction of this article highlights previous research findings that groups typically outperform individuals on tasks which require a single solution. Many researchers conclude that groups perform better due to the social interaction hypothesis, which is the premise that "interaction of the group members not only provides an error-correcting function but facilitates individual thinking and involvement in the task." This study analyzed social context by comparing ad hoc groups with those who have a history of social interaction. The subjects were 137 junior and senior cadets at the U.S. Air Force Academy who were placed into six-person groups.

The resulting data offers evidence that decision quality "is positively related to the amount and quality of interaction among the group members." Also, groups seemed to follow a fairly consistent problem-solving process, which involved exchanging relevant information pertaining to the problem, proposing one or more possible solutions, discussing the possible solutions, and finally, reaching a final decision based on some variant of voting.

- Thorsden, Marvin and Gary Klein. "Cognitive features of team decision making" *IEEE*, 1990. [23]

This article was a summary of work in progress in the development of a team decision modeling tool called the Recognition-Primed Decision (RPD) Model by Klein. Through previous research, the authors found that decision makers in individual and team settings typically evaluate options serially (vs. concurrent) until an acceptable solution is found. One study discussed found that "26 of 27 decisions identified in a five-hour Army battalion-level planning exercise followed a serial format." The RPD is designed to aid team decision making in operational settings. On that note, the article discusses important considerations for the design of such a DSS.

The primary finding was that teams are challenged in balancing the increased coordination necessary when dealing with team members (versus individually) and the increased resources available by having the wider base of knowledge and skills that comes with working on a team. This balancing act impacts the process by which teams make decisions. Whether or not the situation is novel or familiar, the level of interaction and familiarity team members have with one another, and the types of errors made are all related to the balance between coordination and resource utilization. The RPD model aims at being a dynamic model which can adapt to the dynamic decision making inherent in the majority operational settings.

- Cook, Malcolm, L. Elder, and G. Ward, 1997. "Decision Making, Planning, and Teams." *Institution of Electrical Engineers*, Vol. 5, pp. 1-22. [3]

This article is a very broad examination of group and individual decision making. The discussion focuses on past research involving team and individual decision making success and failures, electronic system involvement in decision making, consideration of biases, and user planning and mental models while interacting with such systems as well as other people. The article discusses many different examples in support of ideas stated in the article (i.e. Three mile island, the KLM 747 accident at Tenerife, and the USS Vincennes incident).

Results discussed in the article which are relevant to this research include:

- The role of group members in contributing to effective decisions even though the team leader may have the largest impact on the final decision.
- Factors such as team member personalities, workload, and stress, along with communication, interact with each other and strongly influence decision making performance.
- The overall importance communication effectiveness and methodology can play in effective versus faulty decision making. Flawed decision making is often the result of poor communication techniques of the decision maker.
- How error detection is a factor of each individual person involved in the group (and their respective error detection strategies) as well as how the design of an automated system can aid in detecting errors.
- The level of team cohesion and temporal urgency can significantly impact team decision making processes and effectiveness. More cohesive teams devoted more

time to planning and exhibited more effective task-relevant communication strategies than teams with lower levels of cohesion.

- When workload is high, people find it difficult to consider alternatives or evaluate alternative strategies. In such situations, while working within a team structure, each team member relies solely "on his own interpretation of the situation" due to the lack of confirmation with team members and making revisions.
- Automated planning systems are often ineffective in communicating intent or rationale, so users are cautious about how correct the system may be, which potentially results in low levels of system trust. Therefore, co-operative systems designed with user interaction and cognitive processes in mind may improve human-system interaction and overall decision quality.
- Facilitating good situational awareness is an important consideration for decision making and planning. "Mature decision making relies on good SA and good SA needs effective communication."

➢ Jin, Victoria and Alexander Levis. "Compensatory Behavior in Team Decision Making" *IEEE,* 1990. [7]

This study analyzed team and individual performance under stress. The team performance aspect was set up using two types of organizational structures, hierarchical and parallel. The task involved individuals or 3-member teams working together to detect enemy aircraft and deploy Navy intercept aircraft to defend an air battle environment. One interesting note is that the teams were actually made up of two human decision makers (DM) and one computer.

The key results were 1) team performance was "more stable" when time was decreased, as "team work reduces the effect of individual differences on performance" and 2) hierarchical organization structures performed better than parallel. This result was due to a more effective level of team interaction and communication than that seen in parallel structures. This result is especially relevant in research analyzing military operations as the military type of command structure is hierarchical in nature. In terms of the time pressure variable, time available varied from 30 to 120 minutes, which in some respects is a very large amount of time for military decisions to be made. Although there is a level of time pressure as you allow 30 or 40 minutes to perform a task instead of 100 or 120, additional research could analyze decision making in scenarios where the DM has very limited time available, such as 2-10 minutes.

- Lehner, Paul et al. "Cognitive Biases and Time Stress in Team Decision Making" *IEEE*, 1997. [10]

The effects of time stress were studied in this research with regards to how command and control team's decision making performance varied and if they became susceptible to heuristic bias as a result. Two-person teams performed a military-related scenario in which their task was to defend a battle group from incoming aerial attacks by assessing the type of aircraft associated with certain radar depictions. There were 11 two-person teams in the experiment. Time stress was applied by decreasing the amount of response time (from 45 seconds down to 20 seconds) as aircraft speed increased (from 400 mph to 900 mph).

Results showed that the impact of time stress caused subjects to abandon their training and implement less effective strategies (i.e. decreasing communication, ignoring other DM advice), although this was not what the researchers expected. It was hypothesized that performance would degrade, but statistically it did not. However, the teams were only able to "reduce cognitive workload by simplifying the decision procedure" in a non-optimal fashion.

Influences on Decision Making

- Tjosvold, Dean and Richard Field. "Effects of social context on consensus and majority vote decision making" *The Academy of Management Journal*, 1983. [24]

This study looked at group decision making with respect to social context (competitive or cooperative) and decision strategy (consensus or majority vote). Participants were 114 undergraduate students who were randomly placed into five-person groups. The study consisted of three phases; the first was individual decision making, the second was the forming of groups and reading the group instructions, and the third phase consisted of the group members discussing the issue and making a decision (by either consensus or majority vote). Each group had up to 60 minutes to complete their task. The researchers collected data regarding subject's attitudes towards the group by administering a questionnaire at the end of the task.

Although there were no significant differences in performance across any groups, they found that groups in the consensus condition were significantly more committed to their decisions. Also, cooperative groups showed a greater level of understanding with regards to the problem and they also came to a decision significantly faster than competitive groups did. These results "suggest that a cooperative context encourages learning" and the idea that consensus decision making takes more time and can be frustrating seems to only be true when members "try to compete and outdo one another" instead of working towards mutual benefits.

- Trull, Samuel. "Some Factors Involved in Determining Total Decision Success" *Management Science*, 1966. [25]

Trull investigated 100 case examples from five domains (including the military) in order to gain information regarding successful decision making processes. The case studies generally revealed key variables which seemed to influence decisions in a positive manner. Trull formulated a model comprised of these factors which is shown as follows:

Decision Success = f(Decision Quality) + f(Implementation) where:

Implementation = f(avoid conflict of interest)(risk-reward factor)(degree of understanding)

Pertaining to our research interests, these implementation factors are apparent in various aspects of military decision making. When different organizations (Army, Air Force, Navy or wing, squadron, flight) are involved, there are potential conflicts of interest based on the particular parties involved. These "joint" operations, involving personnel from more than one military service, is typically of today's military operating environment. Also, when dealing with situations the military is involved in and the potentially devastating impact military operations may have, a decision maker must consider the risks and rewards related to any decision that will be made.

Additionally, Trull highlighted a significant area which he noted was often neglected in the case studies he analyzed. Although this literature source is somewhat dated, he did conclude that "the current literature tends to neglect time as an important variable." One of the areas of interest he highlights for future research is that of analyzing the impact of time on the decision making process and its role in decision success.

- Cohen, M. S., Freeman, J.T. and Wolf, S. "Metarecognition in Time-Stressed Decision Making: Recognizing, Critiquing, and Correcting." *Human Factors*, 1996. [2]

The capability to make a good decision can depend on many factors, one of which is the skill level of the decision maker. Cohen, Freeman, and Wolf point out the fact that problem-solving research using a recognition / metacognition (R/M) model shows evidence where expert subjects "are more skilled than novices in critiquing and correcting" information to properly solve a problem. Their research domain was naval tactical decision making. Much like Army battlefield planning, once the situation is analyzed a decision must be made to maneuver assets accordingly. Additional complexity is inherent in the situation due to the dynamic characteristics of military environments. These situational dynamics demand military strategies be multidimensional and unique. Experienced decision makers resolve uncertainty, evaluate time limits, and weigh possible actions better. Their ability to analyze and develop goals may allow them to come to a decision with increased chance for success. The R/M model explains experienced decision makers abilities to handle uncertainty and exploit their experience in a given domain by constructing visual, concrete models. These models lead to improved problem-solving more than abstract strategies because decision makers can manipulate and analyze a concrete situation better due to past experience.

- Lee, John D. and Neville Moray. "Trust, Self-confidence, and Operators' Adaptation to Automation." *International Journal of Human-Computer Studies,* 1994. [12]

Lee and Moray found that operators' utilization of automation depended not only on overall trust in the system but in the operators' perceived ability to control the system. This perceived ability can depend strongly on the person's experience using the system, as Bisantz and Seong found in a 2001 study [1]. It seems logical to deduce the more someone works with a system the more confident and comfortable they should be in understanding the system's capabilities. Self-confidence is another key factor in determining user's perceived ability to control an automated system. Lee and Moray found the combination of trust and self-confidence predicted an operator's strategy when working with automated systems. The same study showed the influence of confidence on automation use when trust exceeds confidence, automation is used. When the opposite holds true, manual operation is preferred. Reliance on automation can be skewed when dealing with overconfidence on behalf of the operator. People are more often overconfident in their own knowledge and abilities but users can also be overconfident in automated system capabilities and accuracy.

3.2 - Synthesis of the Literature

A review of DSS literature has shown how DSS use typically improves performance for individuals and teams, but not in all situations. The benefits of using a DSS have been shown in research involving a variety of domains, from military applications to marketing and economic business situations. However, there are many questions which remain to be answered or at the very least, need clarification as past research has led to some conflicting results.

More research is necessary analyzing the types of decisions (i.e. choosing their own "manual" COAs or the computer-recommended COAs) people make when using a DSS that provides potential problem solutions, particularly when the DSS solutions are brittle or questionable. This issue was addressed in research for *individuals* working with Weasel, where it was found that users may have a propensity of choosing flawed Weasel solutions when brittleness is demonstrated by the DSS [9]. We have yet to address decision type in terms of whether teams (versus individuals) choose solutions generated by the automated DSS tool or their own solutions, as well as what factors influence their decisions (i.e. team expertise, trust in DSS, time impacts, etc).

Research analyzing the impact of DSS brittleness is necessary for a couple of reasons. One, given the impact of decisions in critical domains such as military operations, the influence brittleness has on user behavior is vitally important. Also, as most DSS tools will at some point demonstrate brittleness with regards to the quality of recommendations the tool makes, how that brittle behavior is perceived and addressed by users is crucial to understand.

Analyzing the types of decisions people make when working with automated systems requires an important understanding of the level of trust that users have in the system being operated and their overall trust in automation. Multiple authors have found various factors that lead to trust in automation, including automation reliability [11], operator attitudes [13], workload and situational risk [16]. However, the same factors have not been consistently found in all studies. Parasuraman [15] adeptly states the potential subjectivity of trust in automation by concluding the results of studies “suggest that different people employ different strategies when making automation use decisions” and those people “are influenced by different considerations.”

Parasuraman and Riley [15] highlight the importance of deciding to use (or not use) automation, where the decision “can be one of the most important decisions a human operator can make, particularly in time-critical situations.” In most instances involving complex systems, the interaction between automation and human user must be handled with care considering the stakes at hand, and if the system is known to potentially exhibit brittleness, the interaction between automated system and human user is even more crucial.

Trust in automation literature has highlighted many valuable findings to date, but the majority of the research has focused on individual trust and confidence. Further research analyzing trust and confidence in automation while working within the structure of a team versus a person’s individual capabilities may provide additional insight into the type of decisions people make when part of a team. One important general result typically found when groups are working with highly technical systems is that the automation often removes a significant portion of the procedural tasks, and “the group becomes more concerned with ill-defined tasks, such as planning, learning from experience, handling the unexpected" [8]. This relates closely to our task setup as planning is a key component to deriving and deciding on courses of action.

As discussed above, the power of automation is great but that does not infer that the decision to use or abide by automated aids' recommendations is always correct. This research offered teams the opportunity to decide between automated actions, team-generated actions, or a combination of both. In real world operations the same situation often arises. Airline pilots often have to choose whether to follow a computer recommended flight plan or a route deviation generated by a crew member such as the co-pilot or navigator. Military commanders are given multiple pieces of intelligence data from advanced technological systems as well as information from troops who are in the field seeing battlefield developments in real-time with their own eyes. These sources of intelligence don't always concur with each other one hundred percent. The challenging task of the commander or the entire unit is to decipher which intelligence they feel is best, then plan and act accordingly.

Relative to other complex systems or organisms, humans are adept at taking complex situations involving multiple factors dependent highly on situational factors and accounting for the variables involved. It can be difficult for a computer or other form of automation to do the same. The goal of DSS and specifically the DSS simulation used in this research is to combine the judgmental capabilities of humans with the technological power of computers to create a more powerful evaluation tool than either component could offer by itself.

The capability to make a good decision can depend on many factors, one of which is the skill level of the decision maker. Cohen, Freeman, and Wolf [2] point out the fact that problem-solving research using a recognition / metacognition (R/M) model shows evidence where expert subjects "are more skilled than novices in critiquing and correcting" information to properly solve a problem. Their research domain was naval tactical decision making. Much like Army battlefield planning, once the situation is analyzed a decision must be made to maneuver assets accordingly. Additional complexity is inherent in the situation due to the dynamic characteristics of military environments. These situational dynamics demand military strategies be multidimensional and unique. Experienced decision makers resolve uncertainty, evaluate time limits, and weigh possible actions better [2]. Their ability to analyze and develop goals may allow them to come to a decision with increased chance for success. The R/M model explains experienced decision makers abilities to handle uncertainty and exploit their experience in a given domain by constructing visual, concrete models [2]. These models lead to improved problem-solving more than abstract strategies because decision makers can manipulate and analyze a concrete situation better due to past experience.

How information is presented to people of all skill levels is often as important as what information is presented. Humans are prone to various types of bias in many aspects of life and decision making is no different. Recency effects, representativeness, and availability heuristics are just a few of the types of bias that may affect decision making. Previous research [19] found that presenting an automated recommendation early in a problem solving analysis can significantly impact the decision making process. This impacts the overall situational assessment and evaluation of alternatives. Similar to the work by Smith, McCoy, and Layton, this work analyzed user decisions versus the order in which potential courses of action were presented and developed. A type of decision inertia may be evident where the current use of automation depended on previous use of automation [12]. A situation with ambiguity leaves the decision susceptible to human factors such as an individual's expectations and motivations [22]. It is useful for researchers to recognize and identify any types of bias that may be evident in a subject's decision making process.

The research discussed above appropriately serves as the basis and motivation for the research questions analyzed in this study. Important questions remain to be answered, specifically in relation to Weasel, with regard to team behavior and performance when utilizing a DSS and what potential relationships may exist between user attitudes and their consequent decision making behavior and performance.

3.3 – Research Questions

Based on the literature and interests of the researcher(s) involved in this study, the following questions were developed to analyze DSS usefulness in the military domain.

1. a. When Weasel exhibits brittle behavior, do some teams choose only Weasel's flawed solution set?
    - Hypothesis 1: Teams will choose the computer's flawed solution set more than their personal set.

    b. Does ECOA quality decline when Weasel exhibits brittle behavior?
    - Hypothesis 2: ECOA quality will not decline when the DSS exhibits brittle behavior.

2. a. Does Weasel help teams overall to produce better quality COAs?

    - Hypothesis 3: Teams will have higher quality courses of action when supported by Weasel than when not supported by the DSS.

    b. Does Weasel help users working in teams more than individuals?

    - Hypothesis 4: Teams supported by DSS use will have higher quality courses of action than individuals.

3. a. Does presentation order impact preference toward computer solutions?

    - Hypothesis 5: Viewing computer-generated solutions first will not lead to an increased preference toward choosing the computer solutions.

    b. Does presentation order impact performance?

    - Hypothesis 6: Higher quality ECOA scores will be developed when viewing computer-generated solutions first.

4. How do user attitudes (with regards to communication, satisfaction, confidence, etc.) relate to performance or decision behavior?

    - Hypothesis 7: Teams who have a greater amount of communication will perform better than teams with less communication.
    - Hypothesis 8: Teams who rate their satisfaction and confidence in decisions higher will perform better than teams with lower satisfaction and confidence ratings.

Chapter 4 - Experiment Description

4.1 - Participants

There were 36 total participants for this research study. All participants were cadets enrolled in the University of Minnesota Detachment 415 Air Force (AF) Reserve Officer Training Corp (ROTC) program during the fall of 2009. The participants were assembled into twelve 3-person teams based on the detachment organization structure during the fall of 2009.

Each three person team consisted of one team commander and two subordinates. The ROTC organizational structure was used to comprise teams where the participants were familiar with one another in that they have worked on military training and operations before. This organizational structure closely replicated the command structure of real-world military organizations. The highest ranking cadet in the ROTC structure was designated as the team commander; the other two participants assumed the roles of subordinates. The team roles remained the same for all problem solving scenarios.

Also, it is important to reiterate that the participants discussed as "individuals" in this report refer to the participants who completed the problem scenarios *individually* in the 2004 study. The individual participants are a different set of participants than the 36 participants described above who completed the *team* aspect of this study in 2009.

4.2 - Evaluators

Evaluation of all COAs was based on rankings and quality scores from two evaluators. One evaluator was an active duty AF major with approximately 24 years of total active duty service time in the military. He was prior-enlisted in the US Army for 10 years as an infantry soldier and has served 14 years in the Air Force as a communications officer. With his Army background and communications expertise, he possessed the necessary skills to understand the tasks involved with this research. The other evaluator was an active duty AF major with approximately 14 years of total active duty service time, all as a navigation officer with extensive familiarity with force planning, placement, and operations.

Both evaluators clearly possessed the necessary skills to properly understand the tasks involved in this research and evaluate the corresponding team and individual courses of actions and decisions.

4.3 - Problem Scenarios

Three scenarios were derived for use in experimental trials. The scenarios generally involved a military operational scenario consisting of a given enemy force size and structure along with a geographic depiction of the battlefield area. The purpose of each scenario was to organize the given enemy forces on the battlefield to depict potential courses of action those enemy forces may engage in. Complete descriptions for scenarios 1, 2, and 3 can be found in Appendix 2. Scenario 3 was used as a practice scenario for all teams upon completion of the training to facilitate familiarity with the task and DSS. Scenarios 1 and 2 were used in the research trials and data from those scenarios comprised the data set from which analysis was conducted.

In order to analyze the effect order of presentation and the ability to make revisions may have on subject decision making, two scenarios were made where the computer generated the same number of ECOAs for each scenario. This eliminates any effect the sheer number of automated options presented may have had on the decision maker. Scenarios 1 and 2 each resulted in 8 automated ECOAs generated by Weasel.

Scenario 1 consisted of five units (2 battalions, 2 platoons, 1 company) defending 2 AAs (white and red). The main defense effort was on AA red. The five units are made up of the three different force sizes, the three defense missions (defend, delay, reserve), and various numbers of subunits comprise each unit. The complete description for scenario 1 is shown in Figure 5.

Scenario 1

You've received intelligence from allied ground troops that enemy forces are orchestrating a massive defense in anticipation of being attacked by your allied forces. You want to identify possible enemy defenses so friendly forces know what they may encounter. You know the following about the enemy's situation:

- ❖ Intention is for forces to defend 2 Avenues of Approach (Axis White is to the north, Axis Red is the southernmost AA).
- ❖ There are a large number of enemy forces, made up of 2 battalions, 2 platoons, and 1 company. Details regarding each unit follow:
  - 1 battalion is committed to defend with 3 motorized subunits.
  - 1 company is committed to defend with 2 armored subunits.
  - 1 battalion is defending in delay with 4 mechanized infantry subunits.
  - 1 platoon is defending in delay with 3 armored subunits.
  - 1 platoon is defending in reserve with 2 motorized subunits.
- ❖ The main effort of defense will be on Axis Red, the southern avenue of approach.
- ❖ Forces can be as deep as line of defensible terrain (LDT) 2.

Figure 5. Problem example – Scenario 1

A terrain map of the geographic setup for scenario 1 follows in Figure 6, with each of the 2 AAs labeled (axis white is to the north of axis red).

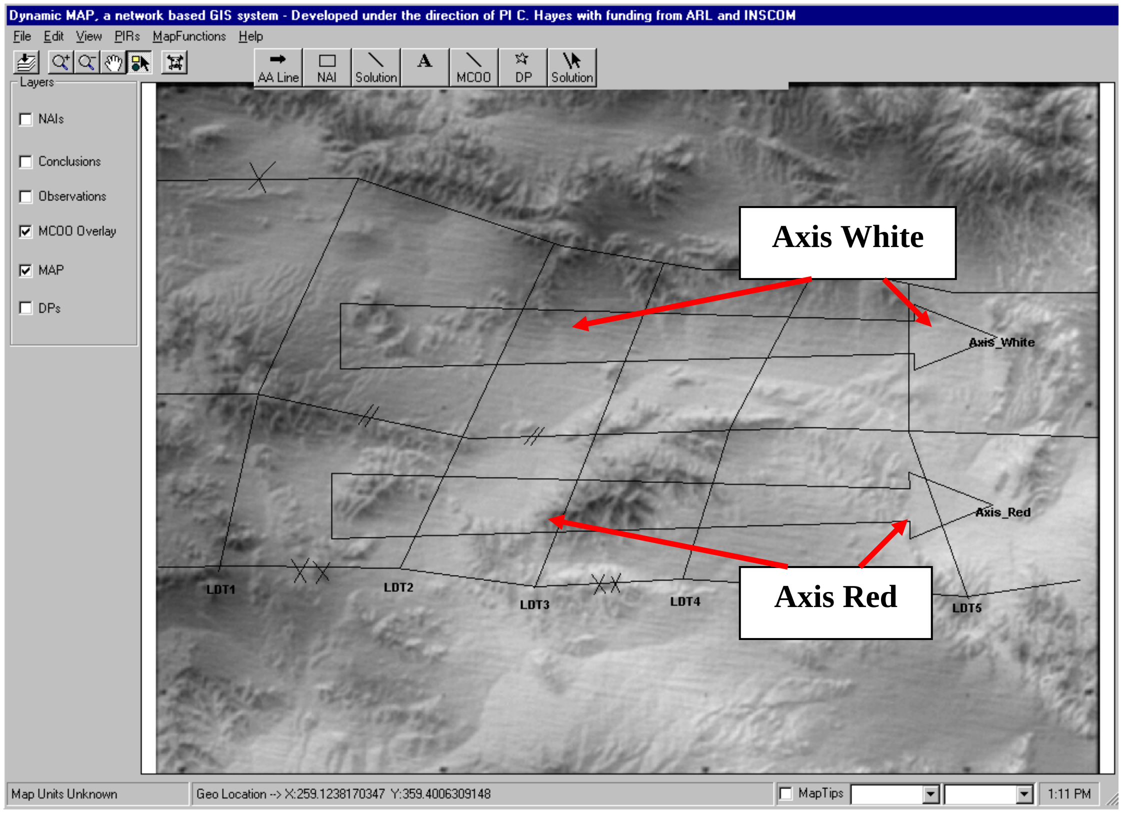


Figure 6. Terrain map of Scenario 1

Scenario 2 was comprised of attacking and defending forces (2 attacking companies, 2 defending battalions, 1 attacking platoon), three AAs were available (Eagle, Crow, Raven), and one additional constraint applied to the situation. The main effort was designated to be on AA crow. Scenario 2 was designed to elicit “brittle” [17] behavior from Weasel, meaning the set of solutions generated by Weasel were flawed in that one AA was always left open in each ECOA. This leaves an opening for counter maneuvers by the opposing force.

Scenario 3 is the most basic of the scenarios and was used as the practice scenario. Intelligence given in the scenario says enemy forces comprised of 4 companies are attacking (2 committed and 2 in reserve), there are two AAs (white and red), and no additional constraints apply to the situation. The main attacking effort was on AA red.

4.4 – Problem Solving Methods

Three problem solving methods were used in the experiment.

Method A:

In method A, teams were first shown a scenario and asked to generated their team or "manual" ECOAs. Next, they were shown the ECOAs generated by Weasel. After reviewing and analysis of the two sets (manual and computer sets), each team made a decision regarding which ECOA(s) were considered the best. For method A problem solving, making revisions to their original manual ECOA set and mixing ECOAs in the final decision was not allowed.

Method B:

In method B, after reading the problem scenario, teams again generated their own ECOAs first. Next, they were shown the ECOAs generated by Weasel. During their review and analysis of the ECOA sets, revisions could be made to their team set. Lastly, each team made their decision regarding which ECOAs were best. Choosing from both ECOA sets was allowed in method B (i.e. choosing a combination of manual and computer ECOAs as "best").

Method C:

In method C, after reading the problem scenario, subjects were first shown the ECOAs from Weasel. Next, they generated their manual ECOAs as well as being allowed to make any revisions deemed necessary. Lastly, they made a decision regarding their ideal ECOA set. For method C, mixing from both sets was allowed.

A lattice [10] experiment design was used, shown in Figure 7. This design counterbalanced any learning effects that may have occurred across the scenarios and afforded the opportunity to analyze effects due to the order of information presented (manual versus automated ECOAs presented to the subject first or second).

Scen. Order	**ECOA Generation Method**					
1/2	A1/B2	A1/C2	B1/C2	B1/A2	C1/A2	C1/B2
2/1	A2/B1	A2/C1	B2/C1	B2/A1	C2/A1	C2/B1
Top row shows the ECOA generation and problem solving method used with each scenario	**A** = manual generation then given computer set of ECOAs (M – A)		**B** = manual ECOA generation then given computer ECOAs then edits (M –A– R)		**C** = shown computer ECOAs then manual ECOA generation then edits (A –M– R)	
Left Column shows the order scenarios are presented in	**1** = scenario 1 **2** = scenario 2					

Figure 7. Experiment design matrix

The design matrix displays the presentation of scenarios for 12 teams. The top row is the ECOA generation and analysis method to be used for the scenarios. The far-left column shows the order in which the team will complete the scenarios.

4.5 – Procedure

Each team completed the experiment with the researcher present at all times. It took approximately 1 ½ hours for each team to complete the experiment. Each participant first read a brief explanation of the study (provided in Appendix 2.6). In this explanation the team's task for experimental trials is described as the following: "The task of subjects will be to evaluate intelligence information, formulate potential enemy courses of action, and analyze courses of action generated by the computer. Considering the given circumstances outlined in the scenario and all relevant information, a decision will then be made to choose the best set of enemy courses of action."

Familiarization training was conducted by the researcher (this took place in the ROTC detachment training room where the problem solving trials were also conducted). Training consisted of a general overview of what a DSS is, and an in-depth explanation and discussion of the "Intel Tool Kit" ECOA and FCOA generator tools, map display, constraint list, and applicable terms and symbols. All participants were permitted to ask any questions at all times during the training. Hard copy printouts of the constraints and acronym and symbology page were provided during the training (refer to appendix 2.8 on page 107 for the constraints printout and figure 4 on page 10 for the symbology used). Teams were given as much time as they needed to study the symbology and constraints before officially starting the scenarios. Also, a practice scenario was conducted for all teams to facilitate familiarity with the task. Prior pilot studies using this specific DSS have shown the use of a practice scenario to be beneficial prior to conducting the actual research trials. The same practice scenario was given to all teams. Training was deemed complete when all pertinent information had been explained and all team members verbally acknowledged they feel comfortable in understanding the experiment, task, and applicable tools.

The necessary materials for each team to complete the scenarios was a scenario instruction page, blank maps for generated team ECOAs, the computer generated COAs, a pen or pencil and the one-page list of constraints. The constraints page was simply a screen capture from the DSS and is included in Appendix 2.8. All materials were in a team/participant booklet and were provided by the researcher.

Before starting the scenarios all participants read the page of instructions shown below in Figure 8. These instructions were also briefed to all team members during the training. Each participant was asked to acknowledge they understood the information described before proceeding.

Scenario Instructions

1. The computer tool finds only the *most likely* ECOAs under the assumptions that have been stated in the scenario concerning enemy resources, intelligence, and likely behavior. It does not necessarily find the *most dangerous* ECOAs.

2. The set of planning rules / constraints listed is *not* necessarily complete.

3. Your goal is to develop a set of ECOAs, which you would give to your commander and to FCOA planners. Soldier's lives may depend on your ability to develop an appropriate set of ECOAs for the given scenario.

4. When you are done generating ECOAs, let the experimenter know.

Figure 8. Scenario instructions

The teams then completed each of the scenarios. Teams were allowed to ask questions pertaining to general understanding during this time. When the teams acknowledge they were finished generating and analyzing ECOAs they were then asked to provide a verbal decision and explanation regarding their choice for the best set of ECOAs for that specific scenario. Participant comments and decisions were recorded by the researcher on a data recording sheet and all teams were videotaped during the completion of the problem solving trials as well. Upon completion of the two scenarios all participants completed two short questionnaires and were then excused from the study.

In summary, the following shows a step-by-step list of what each participant did in this study:

- Upon entering, they were given their subject booklets
- Read brief explanation of the study and signed consent forms
- Completed familiarization training on DSS background, given in-depth explanation of the "Intel Tool Kit" ECOA and FCOA generator tools, map display, constraint list, applicable terms and symbols, and what their task is
- Read scenario instructions and as a team, completed one practice scenario
- Completed two experimental trial scenarios
- Individually, completed questionnaires in back of subject booklet
- Turned in subject booklet and were dismissed

4.6 - Evaluator Procedure

The two evaluators for this study analyzed the quality of the ECOAs generated and the decisions participants made in each scenario. The evaluators performed their analysis completely separate and independent from one another.

Both evaluators completed in-depth familiarization training on the DSS tool prior to evaluating any performance. The Weasel and map display tools were the focus of training, along with constraints, terminology and overall familiarization.

All data given to evaluators was in paper format and all data collected for this study, other than the videotapes of the teams completing the trials, were in paper format as well. The data provided to the evaluators were the participants subject booklets, including the two scenario descriptions (including any notes the participants may have written on the scenario pages), all maps with ECOAs generated by the participants, and the decision made by each team regarding their choice for the best set of ECOAs for the given scenario. Evaluators were also provided the computer generated ECOAs for each scenario, the acronym and symbology sheet used by participants, and the list of constraints which applied in the DSS simulation.

Evaluators were instructed to evaluate performance in each scenario by analyzing the quality of the ECOAs generated (personal and automated) and the decision made by the participant versus the evaluator's ideal solution for the specific scenario. No other specific evaluation criteria was given to the evaluators as the intent of the researchers was to use the military training and operational expertise of the evaluators as valid experience to act as highly capable evaluators for the study. Providing additional "evaluation guidance" may have negatively impacted how these expert evaluators assessed participant solutions and behavior. The validity of the evaluator's ratings was assessed using proven statistical correlation analysis methods to assess if the evaluators indeed provided useful ratings (discussed in greater detail in section 6.1).

Participants were assessed on each scenario based on a 10 point scale and an overall performance ranking and quality score was obtained by averaging the ranks (and scores) from the two evaluators from the two scenarios. In summary, the following shows a step-by-step list of the evaluators' tasks in this study:

- Completed in-depth familiarization training on the "intel tool kit" DSS, focusing on Weasel and map display tools, along with constraints and terminology
- Received hard copy data, consisting of participant subject booklets (including the two scenario descriptions and notes written in the booklet by the participants), all maps with manual ECOAs generated by the teams, and the decision made by each team regarding "best" ECOAs for each scenario.
- Received computer generated ECOAs, DSS constraints and acronym list and symbology sheet for reference during evaluation
- Evaluated participant performance by assigning solution quality scores (based on 10-point scale) to all individual and team solutions.
- Returned materials to researcher

4.7 – Data Recorded

The following data was gathered from each participant during this experiment:

- Participant name and associated participant number
- Date completed experiment
- Participants military experience (years)

- Survey responses
- Team role (commander or subordinate)

The following data was gathered from each team during the experiment:

- Handwritten team ECOAs generated for each scenario
- Number of team ECOAs generated for each scenario
- Number of team ECOAs revised for each scenario (when revisions permitted)
- Decision on best ECOA set for each scenario
- Reason for decision on each scenario
- Verbal comments for each scenario
- Videotaped recording of teams completing each scenario
- Number of verbal communication acts demonstrated from one team member to another during problem solving

At the conclusion of the scenario tasks, each team member was given a survey to complete. The survey consisted of two parts. The first part was the same survey as used in the 2004 research study to acquire subjective ratings regarding participant's trust in automation and their confidence in decisions made with Weasel. The second part of the survey was a perceptual analysis of participant's subjective ratings of 1) aspects of the specific DSS tool used in this study and 2) how they feel their team functioned when completing the tasks. The survey items asked of each subject are shown in Appendix 2.4. The questionnaire rating scales follow the linear numeric design [1] and are commonly used in the United States Air Force for survey administration.

The data sheet used to record information other than the handwritten ECOAs generated by teams and survey responses is shown in Appendix 2.5.

Chapter 5 – Results

5.1 – Participant Solutions

Participant decision choices (computer ECOAs vs. manual ECOAs or both) are shown in Figure 9. The decision labels correspond to the following:

manual = participant(s) chose their own ECOAs which they generated manually,

automated = participant(s) chose the ECOAs automatically generated by Weasel,

both = participants chose a combination of both manual and automated ECOAs

The table depicts the decision and number of ECOAs the participant generated for each scenario. Data is sorted according to participant number, where a "T" before the number indicates a "team." Participants *without a "T" are the individual participants* from 2004.

Subject #	Scenario 1		Scenario 2	
	Decision	# ECOAs	Decision	# ECOAs
1	manual	5	manual	6
2	manual	5	both	5
3	manual	3	manual	6
4	manual	5	automated	4
5	manual	5	both	6
6	both	3	manual	3
7	both	6	both	6
8	automated	8	manual	8
9	automated	4	automated	6
10	both	2	manual	3
11	manual	3	manual	4
12	both	6	manual	4
13	automated	1	manual	3
14	both	4	manual	4
15	manual	2	automated	1
16	both	3	automated	2
17	manual	2	automated	0
18	manual	5	both	6
T1	automated	9	manual	9

T2	automated	7	automated	0
T3	both	3	automated	1
T4	manual	2	automated	6
T5	automated	3	both	5
T6	automated	4	automated	7
T7	manual	2	manual	3
T8	both	22	both	14
T9	automated	0	automated	4
T10	both	4	automated	5
T11	both	5	automated	0
T12	both	2	automated	3

Figure 9. Subject solutions

Shown in Figure 10 is summary data of the decisions made by participants.

# of Participants Choosing Each Option					
Scenario 1		**Scenario 2**		**Total**	
Manual	11	Manual	14	Manual	30
Automated	8	Automated	10	Automated	10
Both	11	Both	6	Both	14
Totals	30		30		60

Figure 10. Summary of subject decisions

5.2 – Evaluator Assessment

5.2.1 - Scenario 1 Scores and Rankings

Figure 11 displays the evaluator solution quality scores for participants in Scenario 1. Included is the score each participant received from the evaluators on the 10-point solution quality rating scale. In evaluating and assigning solution quality scores, each evaluator analyzed the solutions independently from the other evaluator and based participant quality scores according to their perceived quality for the given scenario.

The ranking was derived by the researcher from the solution quality scores collected from the evaluators; the highest score on the 10-point quality rating received the best rank of 1. Rankings continued in this manner until the lowest rank of 30 was assigned to the participant with the lowest quality rating on the 10-point scale.

SCENARIO 1				
	Subject #	**Evaluator 1 Score**	**Evaluator 2 Score**	**Ave Eval 1 & 2 Score**
Individual ECOAs (2004)	1	7.5	6.75	7.125
	2	6	6.5	6.25
	3	5	4.25	4.625
	4	1	1	1
	5	7	8	7.5
	6	8.5	7	7.75
	7	7.75	7.5	7.625
	8	3	1.5	2.25
	9	6	6.75	6.375
	10	4.75	4.5	4.625
	11	2.5	2	2.25
	12	7.5	8.25	7.875
	13	6.5	7.25	6.875
	14	4.5	5.75	5.125
	15	1	1.75	1.375
	16	4.5	4	4.25
	17	1	1.5	1.25
	18	8.75	7.75	8.25
Team ECOAs (2009)	T1	5.25	5	5.125
	T2	5	5.25	5.125
	T3	4	3.5	3.75
	T4	3.5	3.5	3.5
	T5	7.75	8.5	8.125
	T6	5.75	5.5	5.625
	T7	4	3.75	3.875
	T8	8.25	8.75	8.5
	T9	6.25	5.75	6
	T10	8	7	7.5
	T11	7	7.25	7.125
	T12	6.75	6	6.375

Figure 11. Scenario 1 quality scores

Following, in Figure 12, is the corresponding ranks for participants in scenario 1.

SCENARIO 1				
	Subject #	**Evaluator 1 Rank**	**Evaluator 2 Rank**	**Scenario 1 Rank (based on scen 1 ave. score)**
Individual ECOAs (2004)	1	7.5	11.5	9.5
	2	14.5	13	14
	3	18.5	21	20.5
	4	29	30	30
	5	9.5	4	7.5
	6	2	9.5	5
	7	5.5	6	6
	8	26	28.5	26.5
	9	14.5	11.5	12.5
	10	20	20	20.5
	11	27	26	26.5
	12	7.5	3	4
	13	12	7.5	11
	14	21.5	15.5	18
	15	29	27	28
	16	21.5	22	22
	17	29	28.5	29
	18	1	5	2
Team ECOAs (2009)	T1	17	19	18
	T2	18.5	18	18
	T3	23.5	24.5	24
	T4	25	24.5	25
	T5	5.5	2	3
	T6	16	17	16
	T7	23.5	23	23
	T8	3	1	1
	T9	13	15.5	15
	T10	4	9.5	7.5
	T11	9.5	7.5	9.5
	T12	11	14	12.5

Figure 12. Scenario 1 rankings

5.2.2 - Scenario 2 Scores and Rankings

Figure 13 displays the solution quality scores for participants in scenario 2.

SCENARIO 2				
	Subject #	**Evaluator 1 Score**	**Evaluator 2 Score**	**Ave Eval 1 & 2 Score**
Individual ECOAs (2004)	1	7.5	7	7.25
	2	7.25	8.25	7.75
	3	6	6.5	6.25
	4	2.75	2.5	2.625
	5	5.5	6	5.75
	6	7.25	6.75	7
	7	2	3	2.5
	8	5	6.25	5.625
	9	3	2.5	2.75
	10	7	5	6
	11	6	6.5	6.25
	12	4	5.5	4.75
	13	5.5	4	4.75
	14	6	4.25	5.125
	15	2.25	1.5	1.875
	16	2.25	1	1.625
	17	1.5	2	1.75
	18	6.5	4.5	5.5
Team ECOAs (2009)	T1	8	7.75	7.875
	T2	2.5	2.25	2.375
	T3	2.25	1	1.625
	T4	3.25	3.5	3.375
	T5	8.5	9	8.75
	T6	2.75	2.25	2.5
	T7	5.5	7.25	6.375
	T8	8.25	8.5	8.375
	T9	3.5	2.75	3.125
	T10	4.25	3.75	4
	T11	3.25	3.5	3.375
	T12	3.75	3.25	3.5

Figure 13. Scenario 2 quality scores

Figure 14 displays the corresponding ranks for participants in scenario 2.

SCENARIO 2				
	Subject #	**Evaluator 1 Rank**	**Evaluator 2 Rank**	**Scenario 2 Rank (based on scen 2 ave. score)**
Individual ECOAs (2004)	1	4	6	5
	2	5.5	3	4
	3	10	8.5	8.5
	4	23.5	23.5	23
	5	13	11	11
	6	5.5	7	6
	7	29	21	24.5
	8	15	10	12
	9	22	23.5	22
	10	7	13	10
	11	10	8.5	8.5
	12	17	12	15.5
	13	13	16	15.5
	14	10	15	14
	15	27	28	27
	16	27	29.5	29.5
	17	30	27	28
	18	8	14	13
Team ECOAs (2009)	T1	3	4	3
	T2	25	25.5	26
	T3	27	29.5	29.5
	T4	20.5	18.5	19.5
	T5	1	1	1
	T6	23.5	25.5	24.5
	T7	13	5	7
	T8	2	2	2
	T9	19	22	21
	T10	16	17	17
	T11	20.5	18.5	19.5
	T12	18	20	18

Figure 14. Scenario 2 rankings

5.3 – Survey Responses

The data collected via the surveys is provided below. The same questions were asked of all participants, however, "individual" participant's responses are only available on the "simulation decision making and trust in automation" survey (completed in 2004). The 36 participants from the "teams" completed the "simulation decision making and trust in automation" survey as well as the "DSS analysis" survey. The surveys can be seen in Appendix 2.4.

Some descriptions regarding the "simulation decision making and trust in automation" survey questions 5 and 9 follow. The "Wargame Hours Per Week" column represents the average number of hours per week the subject responded they play any type of computerized wargame simulation. A 1 = 0 hours, 2 = 1-2 hours, 3 = 3-5 hours, 4 = 6-10 hours, and 5 = more than 10 hours per week. The last column titled "Better Decision" is the subject's opinion on whether a human or computer would make a better or more trustworthy decision in a situation with multiple variables and potential risk involved.

"Simulation Decision Making and Trust in Automation" survey responses:

Subject #	**General Computer Trust**	**Trust in Computer Analysis for Military**	**Trust in Computer vs. Personal Solutions**	**Computer Attitude**	**Wargame Hours Per Week**	**Confidence Identifying ECOAs**	**Confidence in Computer ECOAs**	**Confidence in Decisions**	**Better Decision**
"INDIVIDUAL" Participant Responses									
1	3	3	2	4	2	3	3	4	human
2	4	2	4	5	2	2	3	4	human
3	5	3	0	5	5	5	3	4	computer
4	4	3	3	5	3	4	3	4	human
5	4	4	3	5	3	3	3	3	human

Subject #	General Computer Trust	Trust in Computer Analysis for Military	Trust in Computer vs. Personal Solutions	Computer Attitude	Wargame Hours Per Week	Confidence Identifying ECOAs	Confidence in Computer ECOAs	Confidence in Decisions	Better Decision
6	4	3	4	5	1	3	2	5	human
7	4	2	4	4	3	5	3	5	human
8	3	3	3	2	1	2	2	2	human
9	4	2	4	4	2	3	2	4	human
10	3	2	5	4	3	3	2	4	human
11	4	3	4	4	1	4	4	4	human
12	4	3	4	5	2	4	4	4	human
13	5	4	3	4	3	3	4	3	computer
14	3	4	4	4	1	3	3	3	human
15	3	1	5	4	2	2	3	3	human
16	4	3	4	5	1	3	3	4	human
17	5	4	4	5	2	2	3	3	computer
18	3	2	5	3	2	3	2	4	human
Ave	3.83	2.83	3.61	4.28	2.17	3.17	2.89	3.72	3C / 15H
"TEAM" Participant Responses									
T1	4	3	3	4	4	4	3	5	human
T2	4	3	4	5	2	3	4	5	human
T3	4	3	3	4	1	4	3	5	human
T4	4	3	4	5	5	4	4	5	computer
T5	4	3	4	3	2	5	4	4	human
T6	4	4	3	4	1	2	3	2	human
T7	3	3	3	3	1	1	4	3	human
T8	4	3	5	3	1	2	4	3	human
T9	5	4	2	4	1	4	4	4	computer
T10	4	3	5	4	2	3	3	3	human
T11	4	4	4	4	1	2	4	4	human
T12	3	4	2	5	1	4	5	3	computer
T13	3	3	2	3	2	2	3	4	human
T14	3	3	4	3	2	2	3	4	human
T15	4	3	4	4	1	4	4	4	human
T16	4	4	3	3	2	1	4	3	computer
T17	4	3	3	3	1	3	4	4	human
T18	4	3	4	3	1	2	4	3	computer
T19	5	3	0	4	1	4	2	5	human
T20	4	2	5	3	1	4	3	5	human
T21	3	3	4	4	1	3	4	4	human
T22	3	3	4	3	1	4	3	5	human
T23	3	2	4	3	1	3	2	4	human
T24	4	3	3	3	1	2	3	4	human
T25	4	3	4	3	1	3	3	5	human
T26	1	2	5	3	1	2	2	3	human

Subject #	General Computer Trust	Trust in Computer Analysis for Military	Trust in Computer vs. Personal Solutions	Computer Attitude	Wargame Hours Per Week	Confidence Identifying ECOAs	Confidence in Computer ECOAs	Confidence in Decisions	Better Decision
T27	4	4	3	5	2	4	5	5	human
T28	4	1	5	4	1	4	3	3	human
T29	4	1	4	4	1	2	4	4	human
T30	4	5	2	3	1	3	4	5	human
T31	3	3	4	4	4	3	4	4	human
T32	4	5	3	5	1	3	5	5	human
T33	4	3	4	5	1	4	4	4	human
T34	3	3	4	4	1	4	4	5	human
T35	3	3	5	3	1	2	4	4	human
T36	5	4	3	5	5	4	3	4	computer
Ave	3.72	3.11	3.56	3.75	1.58	3.06	3.58	4.06	6C / 30H

Figure 15. "Simulation decision making and trust in automation" survey responses

"DSS Analysis: ratings for the specific DSS tool used in the study" survey responses:

(Rating scale: 1 = very poor, 2 = poor, 3 = average, 4 = good, 5 = very good)

Subject #	Quality of Displays	Quality of Intel Info	Quality of COAs	Quality of Information Organization	Quality of Appearance or Style	Ease of Use	Quality as a DM tool	Level of Innovation	Confidence in the Tool	Overall Usefulness	Overall Quality
T1	4	4	3	3	2	4	4	2	3	4	3
T2	5	3	4	5	4	3	5	3	3	4	4
T3	5	5	4	5	5	5	4	4	4	3	4
T4	3	4	4	4	3	4	4	4	3	4	4
T5	2	3	4	3	2	3	4	4	3	4	3
T6	3	4	4	3	3	3	4	4	4	4	4
T7	5	4	4	4	5	3	4	4	3	4	5
T8	4	4	4	5	4	4	4	4	4	4	4
T9	3	4	4	4	2	4	4	3	4	5	5
T10	5	4	4	4	4	4	4	4	4	4	4
T11	4	4	4	4	5	5	5	3	4	4	4

Subject #	Quality of Displays	Quality of Intel Info	Quality of COAs	Quality of Information Organization	Quality of Appearance or Style	Ease of Use	Quality as a DM tool	Level of Innovation	Confidence in the Tool	Overall Usefulness	Overall Quality
T12	4	5	4	5	5	5	5	5	5	5	5
T13	3	3	4	4	3	4	4	4	4	4	4
T14	3	3	4	4	3	4	4	4	3	5	4
T15	5	4	5	4	4	3	4	5	4	4	4
T16	4	2	5	3	4	3	5	5	5	5	5
T17	2	3	3	2	3	3	3	2	3	4	3
T18	4	4	4	4	3	4	5	3	4	5	4
T19	2	2	2	2	2	4	2	4	2	3	3
T20	3	4	3	4	4	4	5	4	4	4	3
T21	5	5	4	5	4	4	4	3	3	4	4
T22	3	3	3	4	4	2	2	3	3	3	3
T23	2	4	2	4	4	4	2	2	3	3	3
T24	4	4	4	5	5	4	5	5	4	5	5
T25	4	4	4	4	3	3	4	3	3	3	4
T26	1	4	4	4	3	3	4	3	3	3	3
T27	4	3	5	4	3	4	5	5	5	4	4
T28	3	3	3	3	3	2	2	5	2	4	4
T29	3	4	4	5	4	4	2	4	4	4	4
T30	4	4	5	4	4	4	5	3	4	4	5
T31	3	4	4	4	4	5	4	4	3	4	4
T32	4	5	5	5	5	5	5	4	5	5	5
T33	3	4	3	3	2	2	4	2	4	4	4
T34	4	4	5	4	5	5	5	4	4	5	4
T35	4	4	4	3	4	4	3	4	4	3	3
T36	3	4	4	4	3	3	5	4	4	4	4
Ave	3.53	3.78	3.89	3.92	3.61	3.72	4.00	3.69	3.64	4.03	3.94

Figure 16. "DSS Analysis: specific DSS tool" survey responses

“DSS Analysis: interacting with your team” survey responses:

(Rating scale: 1 = strongly disagree, 2 = somewhat disagree, 3 = neutral,

4 = somewhat agree, 5 = strongly agree)

Subject #	Trusted input of team members	Subs: trusted DM of leader	Preferred involvement in final decision	Preferred leader to make decision	Comm effective when everyone made decision	Comm effective when leader made decision	Better decision when all decide	Better decision when leader dec.	Overall team effective in decision quality	Overall team effective in comm	Overall team effective together
T1	5		5	3	4	4	5	3	5	4	4
T2	5	5	5	3	5	3	5	3	5	4	4
T3	5	5	5	4	5	4	5	4	5	5	5
T4	5		5	5	4	5	5	3	4	5	4
T5	5	5	4	4	5	5	4	4	5	4	5
T6	5	5	5	2	5	5	5	2	5	5	5
T7	5		4	4	5	5	3	2	4	5	5
T8	5	5	4	3	4	3	4	3	4	5	5
T9	5	5	5	4	5	2	5	2	5	5	5
T10	5		5	3	5	3	4	3	3	4	4
T11	4	4	5	2	5	3	5	3	5	4	4
T12	5	5	4	4	5	5	5	4	5	5	5
T13	5		5	5	4	5	4	5	4	5	5
T14	5	5	5	2	5	3	5	3	5	5	5
T15	5	5	5	4	4	5	4	5	5	5	5
T16	5		5	3	4	3	4	3	4	4	4
T17	5	5	4	3	4	4	4	3	4	5	5
T18	4	5	4	3	5	5	5	4	4	5	5
T19	5		5	4	5	4	5	2	5	5	5
T20	5	4	5	1	5	3	5	3	5	5	5
T21	5	4	5	4	5	3	5	3	5	5	5
T22	4		5	5	4	3	3	3	4	4	4
T23	5	5	5	5	5	5	5	3	5	5	5
T24	5	5	3	4	4	3	3	4	4	4	4
T25	4		5	4	4	3	5	3	4	3	3
T26	4	4	5	4	3	3	3	3	4	4	4
T27	5	5	5	1	3	3	5	2	5	5	5
T28	5		4	4	4	3	5	3	5	5	5
T29	4	5	5	2	4	4	3	3	4	5	5
T30	5	5	5	4	5	5	4	4	4	5	5
T31	4		4	4	3	4	3	4	4	3	4
T32	5	5	5	2	5	5	5	3	5	5	5
T33	5	4	5	2	3	5	5	2	5	4	4
T34	5		5	5	5	4	5	4	5	5	5

T35	5	5	4	5	4	4	3	4	4	5	5
T36	4	4	5	4	3	3	5	3	5	3	4
Ave	4.78	4.75	4.69	3.47	4.36	3.86	4.39	3.19	4.53	4.56	4.61

Figure 17. "DSS Analysis: interacting with team" survey responses

Chapter 6 – Analysis

The first analysis step was to see if the evaluators' rankings were consistent with one another and therefore useful? Once the validity of the rankings and quality scores was established, they were used to assess the following questions:

- Does Weasel help teams overall to produce better quality COAs?
- Does Weasel help users working in teams better than individuals?
- When Weasel exhibits brittle behavior, do some teams choose only Weasel's flawed solution set?
- Does ECOA quality decline when Weasel exhibits brittle behavior?
- Does presentation order increase preference toward computer solutions?
- Does order of presentation impact performance?
- How do user attitudes (with regards to communication, satisfaction, confidence, etc.) relate to performance or decision behavior?

The answers to these questions will be used to assess the impacts of using such a system, how Weasel should be used and by whom, and potential future considerations for DSS applications.

6.1 – Are evaluators' rankings consistent?

There were two correlation methods used to assess the level of agreement between the two evaluators. Basic numerical correlation analysis was performed on the solution quality scores and the Spearman rank correlation [27] was used to measure the correlation of rankings. Overall, both correlation methods indicate there is a very high level of agreement between the evaluators.

The correlation r values between the two evaluator's solution quality scores in Scenario 1 was .948 and Scenario 2 was .907, indicating a high level of agreement between the evaluators. Given the typical standard of a correlation of 0.90 or higher indicating a high level of agreement, the experimenter felt justified in trusting the evaluator's solution quality scores.

Quality Score Correlation Values
Scenario 1: r = .948
Scenario 2: r = .907

Figure 18a. Quality score correlation values

Regarding rankings, the Spearman rank correlation value, r_s, is computed using the following formula:

$$r_s = 1 - [6(\sum_{j=1}^{n} d_j^2)] / [n (n^2 - 1)]$$

The rank correlation value can range from -1 to 1, where -1 is perfect disagreement and 1 is perfect agreement between the evaluators. The d is the difference between the assigned ranks from the evaluators, n is the total number of solution alternatives, and j is the j^{th} solution alternative.

Several solutions were identical in quality. Those solutions were assigned the average rank of the two adjacent rankings. An example is in Scenario 1 where two solutions tied for the fifth best ranking according to the first evaluator (team 5 and individual 7). These two solutions each received a rank value of 5.5 derived from (5 + 6)

/ 2 = 5.5. The next solution received a rank value of 7, the next rank value in the sequence.

The Spearman Rank Correlation between the two evaluators in Scenario 1 was .934 and Scenario 2 was .921 and the overall rankings led to a Spearman Rank Correlation of .939. Spearman Rank Correlation calculations are shown in Appendix 3.1. This shows a very high level of agreement between the evaluators, on average. Once again, given the typical standard of a correlation of 0.90 or higher indicating a high level of agreement, the data obtained in this research indicates the evaluators' assessment was consistent and can be utilized for data analysis and interpretation.

Spearman Rank Correlation (r_s) Values
Scenario 1:   r =   .934
Scenario 2:   r =   .921

Figure 18b. Spearman rank correlation values

Provided in Figure 19 are the overall rankings based on the average of the scores each participant received from the two evaluators in each of the scenarios. Included in the table are the participant's group type (individual or team), average score from the evaluators, and overall ranking for the 2 scenarios. The best rank is 1 through the worst ranking of 30.

OVERALL Rankings			
Subject #	**Group Type**	**Average Score**	**Overall Rank**
T5	Team	16.875	1.5
T8	Team	16.875	1.5
6	Individual	14.75	3
1	Individual	14.375	4
2	Individual	14	5
18	Individual	13.75	6
5	Individual	13.25	7
T1	Team	13	8
12	Individual	12.625	9
13	Individual	11.625	10
T10	Team	11.5	11
3	Individual	10.875	12
10	Individual	10.625	13
T11	Team	10.5	14
T7	Team	10.25	15.5
14	Individual	10.25	15.5
7	Individual	10.125	17
T12	Team	9.875	18
T9	Team	9.125	19.5
9	Individual	9.125	19.5
11	Individual	8.5	21
T6	Team	8.125	22
8	Individual	7.875	23
T2	Team	7.5	24
T4	Team	6.875	25
16	Individual	5.875	26
T3	Team	5.375	27
4	Individual	3.625	28
15	Individual	3.25	29
17	Individual	3	30

Figure 19. Overall rankings across all scenarios (1 = best; 30 = worst)

6.2 – Replication of Prior Results for Individuals

Due to the fact that different evaluators provided the solution quality scores for this study, it is worth noting how the data and results from this study (and the resulting evaluator's scores) compare with those from the 2004 study in which individuals completed the problem solving scenarios utilizing Weasel. The following is a summary of this comparison:

- *Do the correlations between the 2004 evaluator scores and 2009 evaluator scores for each scenario agree with each other?*

  To gain further clarity into the level of agreement between the evaluators from the 2004 study and the current study (completed in the fall of 2009), correlation analysis was conducted on the individual participant's solution quality scores for each scenario. The correlation value between evaluator scores for scenario 1 was r = .868 and for scenario 2 the correlation r-value = .901. These values indicate a relatively high level of agreement between the individual's solution quality scores as assessed by both sets of evaluators from the two studies.

- *Does Weasel help individual users overall to produce better quality ECOAs?*

  In 2004, an important result found individual users did indeed produce better ECOAs with the use of Weasel than without (p-value = .018). Based on the current evaluator's solution quality scores, the same result is replicated, where the individuals produced better quality ECOAs when using Weasel than when not using Weasel (p-value = .025). Graphs of the data for each study are shown in Figure 20 and 21, respectively.

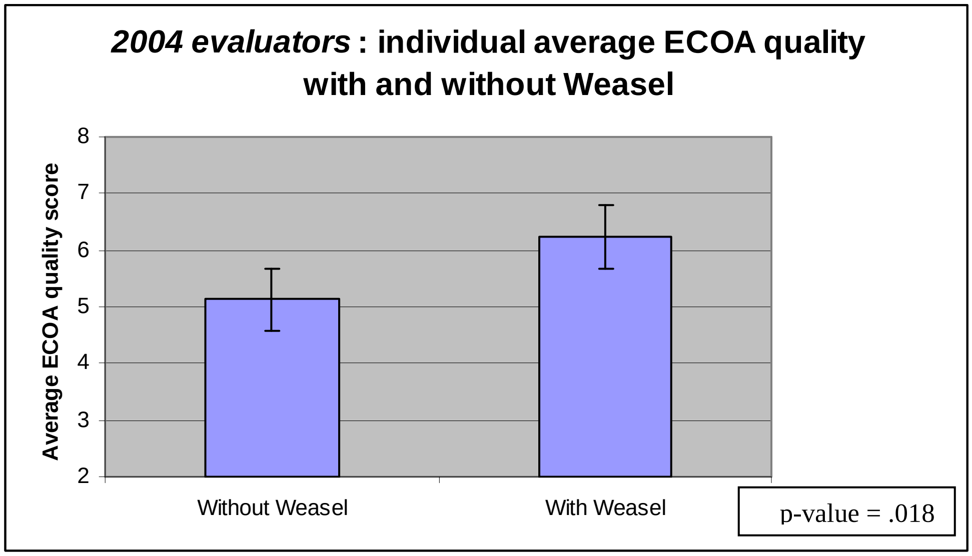


Figure 20. 2004 evaluators: individual solution quality with and without Weasel

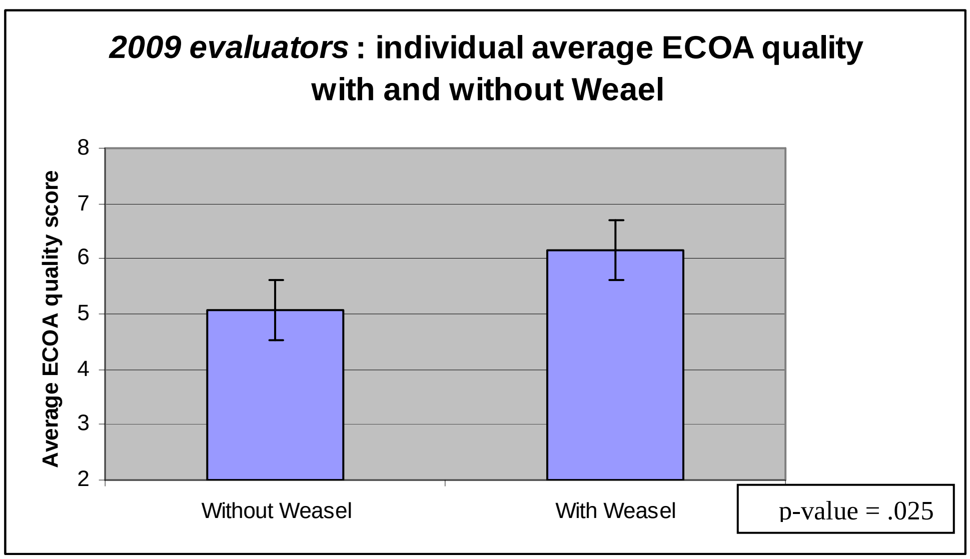


Figure 21. 2009 evaluators: individual solution quality with and without Weasel

➢ *Does ECOA quality decline for individual users when Weasel exhibits brittle behavior?*

The 2004 study showed individuals' ECOA quality scores were not significantly lower in the brittle scenario when using Weasel (p-value = .51). Once again, this result is replicated using the current evaluator scores, with a resulting p-value of .46. Graphs of the results from each study can be seen in Figure 22 and 23 respectively.

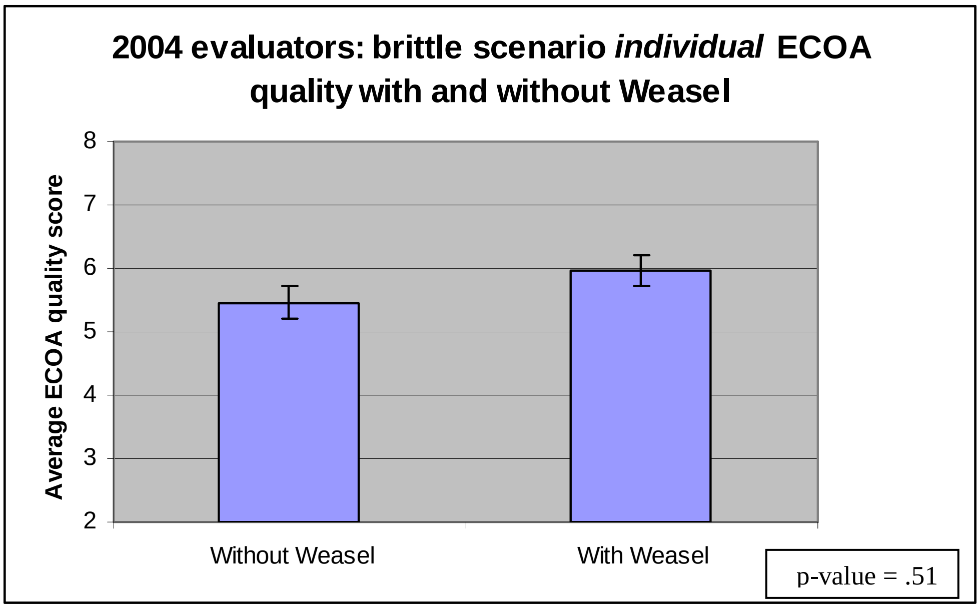


Figure 22. 2004 evaluators: brittle scenario individual solution quality with and without Weasel

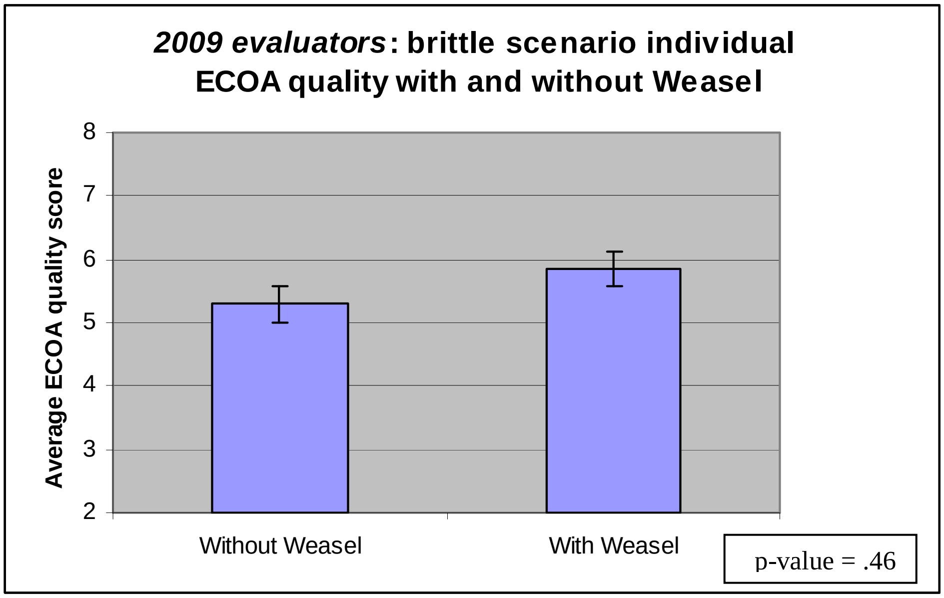


Figure 23. 2009 evaluators: brittle scenario individual solution quality with and without Weasel

- *Does presentation order impact individual performance?*

Viewing the Weasel ECOAs first led to significantly better solution quality scores for individuals in 2004 (p-value = .018). Using the current evaluator's scores, the same result was not replicated; the resulting p-value of .52 indicates no significant differences in individual solution quality scores based on viewing automated ECOAs first or generating manual ECOAs first. This different result may simply be due to the fact that different evaluator's were used in the studies or to a more complicated factor that isn't readily apparent. Graphs of the data for this analysis are shown in Figures 24 and 25, respectively.

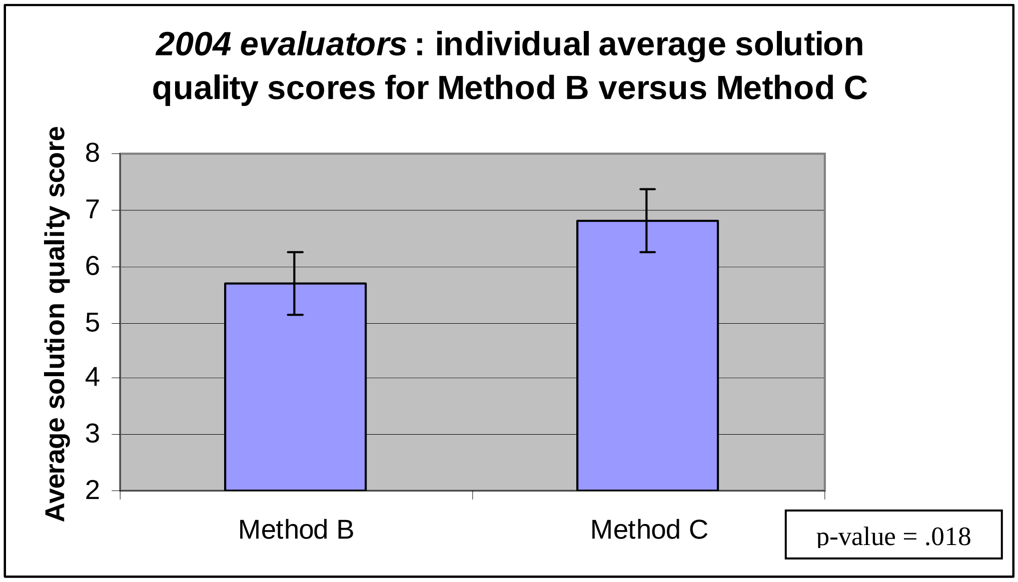


Figure 24. 2004 evaluators: individual average quality scores for Method B vs. Method C

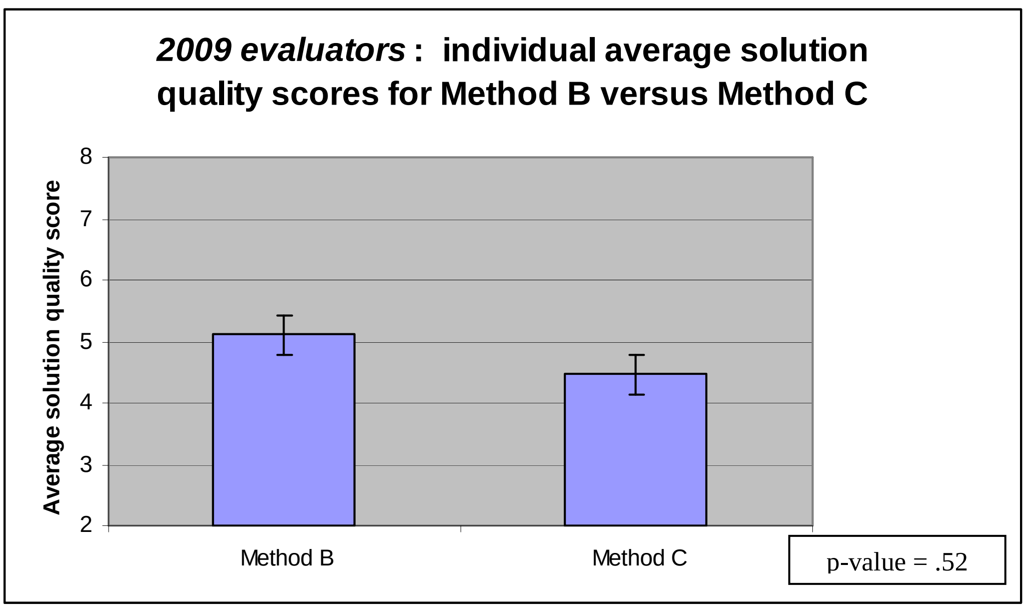


Figure 25. 2009 evaluators: individual average quality scores for Method B vs. Method C

6.3 – When Weasel exhibits brittle behavior, do some teams choose only Weasel's flawed solution set?

Yes, 8 of 12 teams chose only ECOAs from the DSS flawed solution set in scenario 2, while 2 of the 12 groups chose their team ECOAs, and the remaining 2 teams chose a combination of both computer and team ECOAs (and both of these teams purposefully chose ECOAs they generated that covered all three avenues of approach). The resulting ANOVA p-value for this analysis was equal to .011, showing that teams chose the computer's flawed solution set significantly more than their own manual ECOAs or a combination of both ECOA sets. As a reminder, if subjects chose only Weasel's ECOA set, which all showed the enemy leaving one AA uncovered, the decision was considered to be of poor quality. Evaluators felt it was unrealistic to assume the enemy would always leave one AA open. Therefore, the fact that a significant number of teams chose the flawed Weasel ECOA clearly had a direct negative impact on the quality of decisions made by those teams.

This result also confirms the trend evident in the 2004 study where 5 of 18 individual subjects picked only the Weasel ECOA set in the scenario in which Weasel exhibited brittle behavior.

A tabular summary of the results for the "brittle" scenario follows:

Scenario Decision			
	Manual Set	Computer Set	Combined Manual & Computer Set
Teams (p-value = .011)	2/12	**8/12**	2/12
Individuals (p-value = .205)	9/18	**5/18**	4/18

Figure 26. Brittle scenario subject decisions

6.4 – Does ECOA quality decline when Weasel exhibits brittle behavior?

No, ECOA quality and decision quality did not significantly decline when Weasel exhibited brittle behavior. Although the result was not statistically significant, descriptive statistics show that on average, the quality of ECOAs decisions by both individuals and teams declined in the brittle scenario. Considering the result discussed above in section 6.3 where the brittle ECOAs were chosen by many of the participants, this may show a trend and potential large impact regarding the effect the brittleness may impart on user decision quality. This behavior is important in that although almost all teams clearly noticed that one AA was always left open in the Weasel ECOA set and most felt that was low quality in their initial analysis, as the results show a significant number of the teams still eventually chose the flawed solutions. They discussed the open AA in depth, and many teams initially commented that such a maneuver would be "stupid," "very dangerous," and "that doesn't make sense." Yet after discussion and analysis they seemed to still trust the Weasel more than their own intuition.

For individuals, their ECOA quality declined from 5.13 in scenario 1 to 4.73 in scenario 2 (the brittle scenario). For teams, there was a similar result where quality declined from 5.88 in scenario 1 down to 4.60 in scenario 2. These results are also shown below in figure 27.

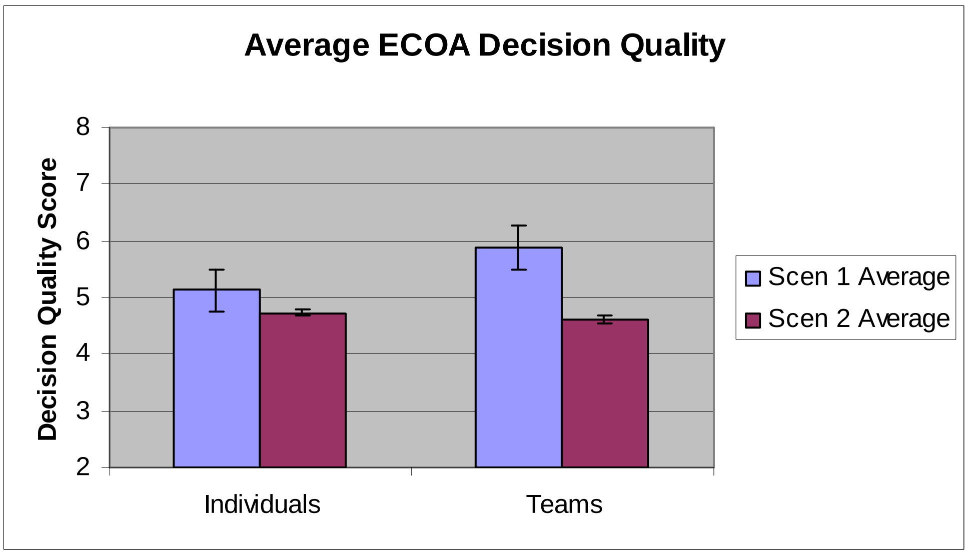


Figure 27. Average ECOA decision quality scores between scenarios 1 and 2

6.5 – Does Weasel help teams overall to produce better quality COAs?

No, the analysis revealed that teams did not produce significantly higher quality ECOAs with the assistance of Weasel than without. The analysis consisted of comparing quality scores for manual ECOAs (Method A and Method B without revisions) versus ECOAs generated with the assistance of Weasel (Method's B and C). The resulting

ANOVA p-value was .50; indicating there was no significant difference in quality when teams were assisted by Weasel. However, when simply comparing average scores, teams did produce slightly higher quality ECOAs with weasel than without (average of 6.34 with Weasel vs. average of 5.69 without Weasel).

This result is different than that found in a related analysis in the 2004 study on whether Weasel helped individual planners to produce better quality COAs. That analysis revealed that users produced significantly higher quality ECOAs with the assistance of Weasel than without. The resulting ANOVA p-value was .018. This may potentially be due to the level of expertise a team may have versus just one individual working alone, or possibly the social dynamics and or complexities of working as a team hinder a team's performance when attempting to utilize a DSS such as Weasel.

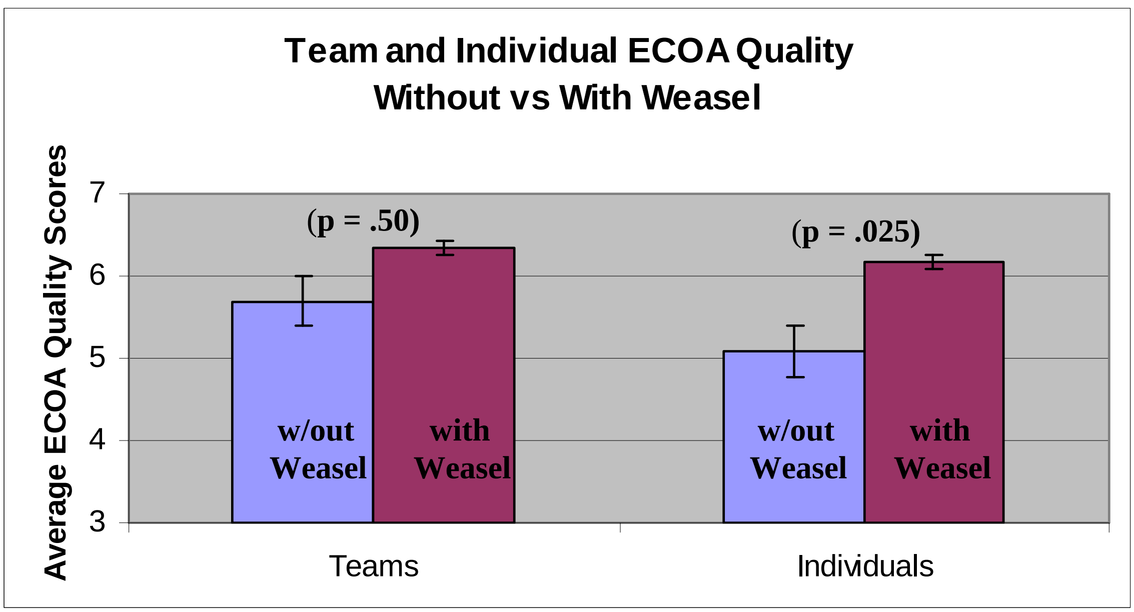


Figure 28. Overall Team and Individual ECOA Quality: Without vs. With Weasel

Another interesting result was that the computer produced higher quality ECOAs than humans in scenario 1 (the "non-brittle" computer solution set), even when the human was assisted by Weasel. Also, as shown in the figure below, the computer performed significantly worse than humans when Weasel exhibited brittle behavior, as in Scenario 2.

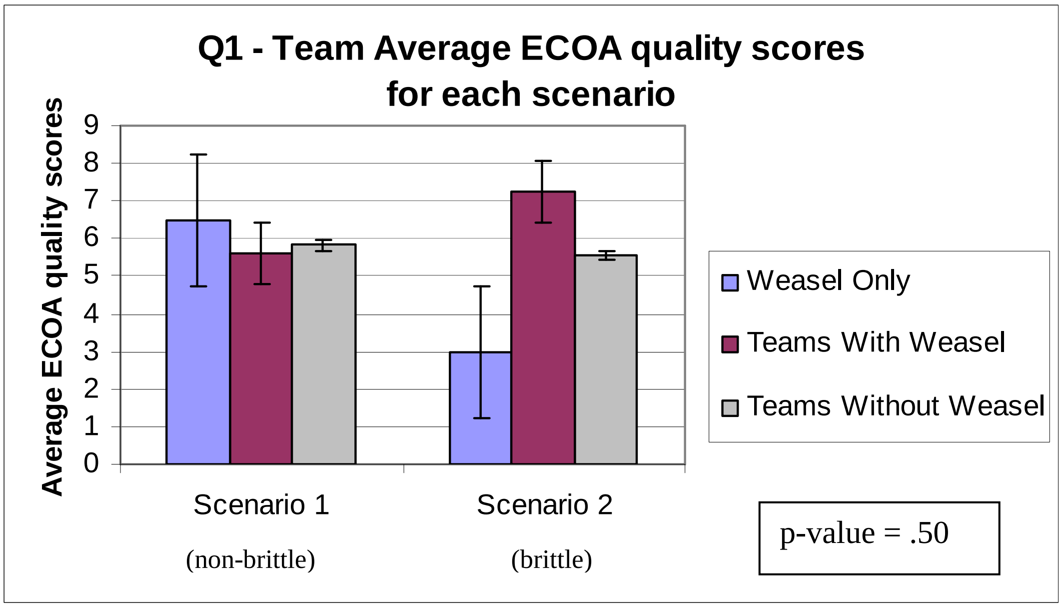


Figure 29. Team average ECOA quality (ECOAs produced manually by teams) for each scenario

6.6 – Does Weasel help users working in teams more than individuals?

No, the analysis revealed that teams did not perform significantly better than individuals with regard to either rank or decision quality. Regarding overall decision quality scores, the ANOVA resulted in a p-value of .658 and the p-value was .943 in terms of overall ranking between teams and individuals. The same type of results held true for each scenario. In scenario 1, the decision quality score p-value was .376 and rank was .569. In scenario 2, the decision quality score was .882 and analysis on differences between team and individual rankings led to a p-value of .916. These numbers clearly indicate that in terms of performance, there was no significant difference between teams and individuals.

6.7 – Does presentation order impact preference toward computer solutions?

There is no significant difference in team choices between viewing automated ECOAs first and manually generating ECOAs first, which is different than the hypothesis stated previously that seeing the computer ECOAs first would create a preference toward choosing the computer ECOA set. Interestingly enough, the results show a trend counter to what may seem intuitive, where 75% of teams actually chose the computer set in Method A, in which they generated teams ECOAs first, versus only 50% in Method C, in which they viewed the computer generated ECOA set first. In fact, in the eight instances when teams generated their manual set of ECOAs first, they always chose either computer ECOAs only or a combination of computer and manual ECOAs.

The analysis for this question compared decision type (choosing manual, automated, or both types of ECOAs) for the scenario in which the automated ECOAs were viewed first (Method C) versus the decision made in which manual ECOAs were generated first (Method A). The p-value was .334. The ANOVA data can be seen in Appendix 3.2.

For all scenarios in which teams viewed the computer generated ECOAs first, only 4 of 8 (50%) teams chose the automated ECOA set as their ideal set. Two teams

(25%) chose their own manual ECOA set as ideal, and two teams (25%) chose a mix of both manual and automated ECOAs.

When teams created manual ECOAs first using problem solving method A, manual ECOA sets were never chosen (0 of 8 instances), automated sets were chosen six times (75%), and a mix of manual and automated ECOAs were chosen in two instances (25%).

p-value = .334		**Presentation Order**	
		View Automated First (Method C)	Generate Manual First (Methods A)
Team Decision	Manual Set	25%	0%
	Automated Set	50%	75%
	Mix of Manual & Automated	25%	25%

Figure 30. Team decision frequency for presentation order (method C vs. A)

A useful result related to this question does show that teams did choose the computer generated ECOAs significantly more often than individuals did. This was true for the overall number of decisions made (p-value = .010) as well as the decisions made in scenario 2, the “brittle” scenario (p-value = .036). The following figure shows a breakdown of the decisions made by teams and individuals.

Team vs Individual Decisions						
	Scenario 1		Scenario 2		Overall	
Decision	Teams	Indiv	Teams	Indiv	Teams	Indiv
Manual	2/12	9/18	2/12	9/18	4/24	**18/36**
Automated	5/12	3/18	**8/12**	5/18	**13/24**	8/36
Both	5/12	6/18	2/12	4/18	7/24	10/36
Analysis Results	No significant differences		Teams chose automated more (p-value = .036)		- Teams chose automated more (p-value = .010) - Individuals chose manual more (p-value = .008)	

Figure 31. Decision-Type Summary

6.8 – Does presentation order impact performance?

Analysis showed teams who viewed automated ECOAs first did not perform any better than those that generated manual ECOAs first. The resulting p-value was .389. As mentioned previously in section 6.2, the same result holds true for individual users as well (p-value = .52).

To answer this question, analysis was performed comparing team ECOA quality between presentation method B and method C across all scenarios. Method B was manual ECOA generation followed by viewing automated ECOAs, with editing and mix and match decision allowed. Method C was viewing automated ECOAs first, and then generating manual ECOAs, with editing and mix and match allowed.

6.9 – How do user attitudes (regarding communication, satisfaction, confidence, etc.) relate to performance or decision behavior?

Responses to relevant questions can often provide useful insight into participant behavior and performance. Analyses (ANOVAs) were conducted on various questions asked of research participants upon completion of their experimental trials. Discussion will be given here to questionnaire items that led to statistically significant results.

Individuals versus Teams

"Simulation Decision Making and Trust in Automation" Survey

There were three items with statistically significant differences between participants who worked as individuals and those who were part of a team.

- *Question 4: What rating best represents your attitude about computers?*

Participants rated their attitudes about computers (scale from 1 = completely dislike computers to 5 = completely like computers), and the results showed individual participants had a significantly higher ratings overall than team participants did (4.28 average versus 3.75), with a resulting p-value of .024.

- *Question 5: On average, how many hours per week do you play any type of computerized wargame simulation? (scale of 1 = 0 hours/week up to 5 = more than 10 hours per week)*

This question showed a significantly higher average for individuals than team participants (p-value = .049).

- *Question 7: Rate your confidence in the computer identifying important enemy courses of action in this experiment. (scale of 1 = low confidence to 5 = high confidence)*

The team participants had a significantly higher confidence rating in the computer than individuals (p-value = .002). This result may demonstrate one reason for the higher propensity of teams to choose computer ECOAs more than individuals.

<u>Leaders versus Subordinates</u>

There were five questions in which leaders and subordinates demonstrated significantly different responses. These items are discussed below.

<u>*"DSS Analysis: ratings for the specific DSS tool used in the study" survey*</u>

(rating scale: 1 = very poor, 2 = poor, 3 = average, 4 = good, 5 = very good)

- *Question 2: Quality of intelligence information*

  This question showed subordinates rated the quality of the intelligence significantly higher than team leaders (p-value = .031).

- *Question 9: Confidence in the tool*

  Subordinates had significantly higher confidence ratings than team leaders (p-value = .028) with regards to their confidence in the Weasel DSS.

"DSS Analysis: interacting with your team" survey

(rating scale: 1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, 5 = strongly agree)

- *Question 4: I preferred the team leader make the final decision*

With a p-value of .019, the results showed team leaders preferred to make the final decision significantly more than subordinates wanted the leaders to make the decision.

- *Question 9: Overall, our team was effective in the quality of our decisions*

Subordinates agreed with the statement significantly more than team leaders (p-value = .033) regarding the overall quality of the decisions the team made during the problem solving scenarios.

- *Question 11: Overall, our team was effective in working together*

Subordinates had significantly higher ratings than team leaders (p-value = .029) regarding the overall effectiveness of the team.

Responses to item 9 and 11 may be due to leaders being more critical of their performance and behavior as well as that of their team. Team leaders feel the responsibility for their team performance and their perspective of the team's performance may be different than that of team subordinates.

The results stated above show interesting relationships and potential factors that may impact user behavior and performance. Social dynamics and individual team member attitudes and perceptions may be influencing team behavior. Also, the complexities that often come when working as a team and trying to solve relatively complex problems may influence the mindset of team members versus individual users who are problem solving and working with a DSS such as Weasel by themselves. These results will be discussed further in the "interpretation of the results" and "conclusions" section of the paper.

On an additional note, it is worth pointing out the fact that there were no statistically significant differences between top performing teams and bottom performing teams on any of the survey items.

6.10 – Additional Analyses

Two additional analyses were performed based on areas of interest pertaining to this topic of study, specifically regarding whether a team's amount of communication may impact their performance and if there may be a relationship between the number of ECOAs a user generates and their overall solution quality.

From the videotaped recordings of team experimental trials, the total number of acts of communication (communication relevant to the problem scenarios) between team members was recorded. This data was used to analyze potential relationships between levels of team interaction/communication and team performance. An ANOVA statistical analysis revealed a significant difference in the amount of communication between the four top-performing teams and the four lowest-performing teams (performance analyzed with regard to rank). Higher-ranked teams had a significantly greater amount of communication than lower-ranked teams (p-value = .017), averaging 139.25 total communication acts versus only 82 acts of communication for lower-ranking teams. The specific data pertaining to this analysis is shown in Figure's 32 and 33, and the ANOVA output can be seen in appendix 3.2.

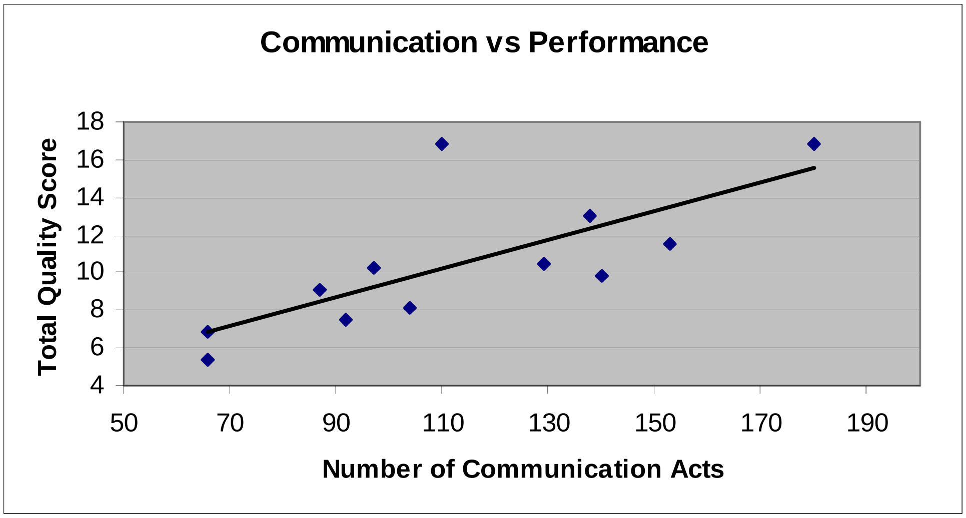


Figure 32. Team Communication vs. Performance

Team Communication vs Performance		
	Top 4 Teams	**Bottom 4 Teams**
Total "Acts" of Communication	180	104
	110	92
	138	66
	129	66
Average (p-value = .017)	139.25	82

Figure 33. Communication and Performance: Top Performing Teams vs. Bottom Teams

An analysis was also performed to determine if a relationship may exist between the number of ECOAs generated by a decision maker (regardless of whether or not they chose any of those ECOAs) and the overall solution quality. A linear regression analysis was performed for each scenario as well as for the total number of ECOAs generated for all scenarios. The analysis revealed a statistically significant positive relationship between the number of ECOAs generated and solution quality scores on scenario 2 (correlation coefficient = .74 and corresponding p-value = .001) and overall (correlation coefficient = .832 and corresponding p-value = .006), but not on scenario 1 (correlation coefficient = .496 and corresponding p-value = .124). Due to a large outlier data point (from team 8) in the total ECOAs generated, an additional regression analysis was performed with the outlier removed. This analysis also revealed a significant positive relationship between the number of total ECOAs generated and a team's total solution quality score, with a p-value = .0377. Based on this data, it seems there may be a positive relationship between generating more ECOAs and overall solution quality. This may be due to the fact that simply brainstorming and generating more ECOA options leads to more and potentially better ECOAs to pick from as well as better analysis overall in deciding which ECOA(s) are best.

A plot of the results for overall data and scenario 2 data can be seen in Figure's 34-36. The specific regression output for this analysis can be seen in appendix 3.3.

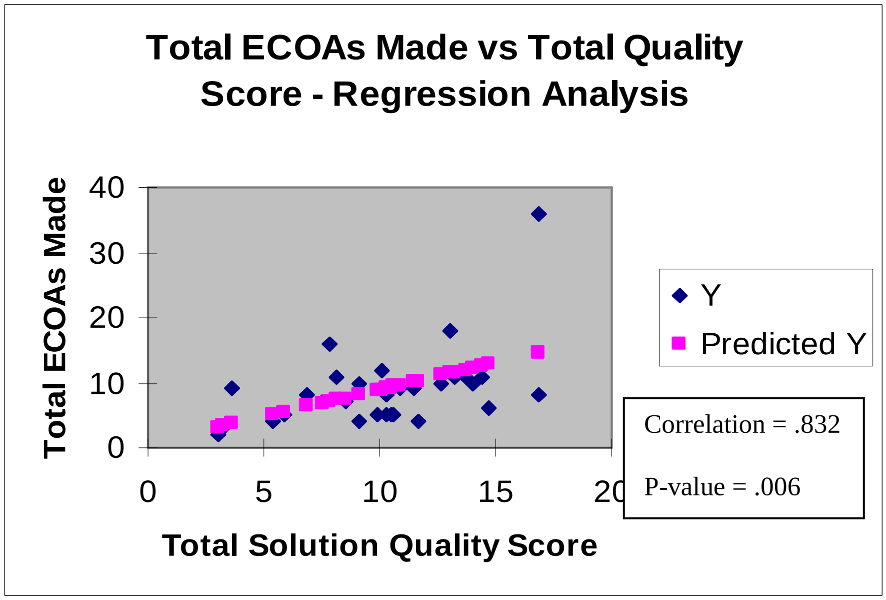


Figure 34. Regression Analysis: Total ECOAs vs Total Quality Score

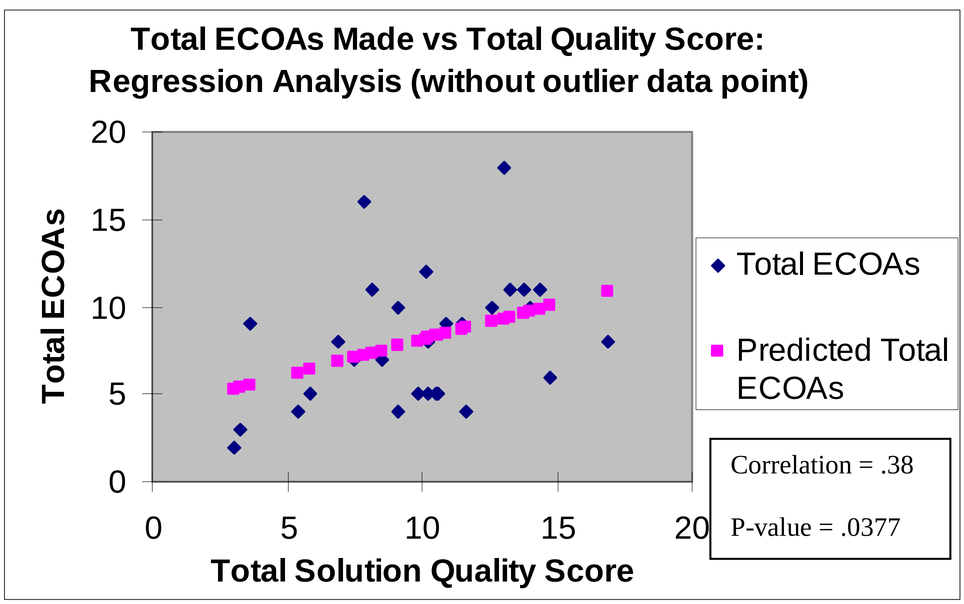


Figure 35. Regression Analysis: Total ECOAs vs Total Quality Score (WITHOUT OUTLIER)

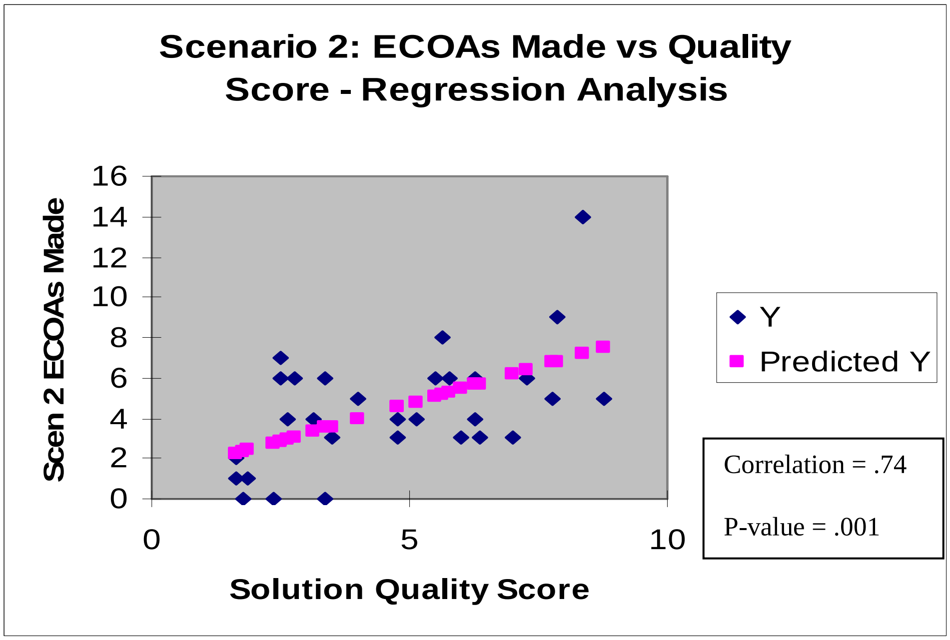


Figure 36. Regression Analysis: Scenario 2 ECOAs Made vs Quality Score

6.11 – Participant Comments

The 36 participants for this study were asked to provide comments and recommendations regarding the DSS used in this study, Weasel, as well their thoughts pertaining to effectiveness of decision making, problem solving, and any other aspects of interacting as a team. The following is a representative sample of comments received via the questionnaires.

Topic	Comment	Additional Information
Weasel Specific Comments	"This system is a good way for a commander to get all of the information. Then lead into action on the best possible plan."	Team leader
	"The usefulness of the DSS was efficient in getting an outside source on tactics."	Subordinate
	"I liked the scenario better where we had to generate our own ideas first…and then look at the computer results."	Subordinate
	"I think the computer results were a good 'check and balance' to help us see if we were on the right track or completely off."	Subordinate

Weasel Specific Comments	"Easy to use and pretty effective."	Subordinate, no DSS experience
	"It was a useful program but may lack elements of surprise. Good for gathering a census of what enemies are or could be doing."	Subordinate
	"It recognized COAs that our team would not have otherwise identified."	Subordinate
	"I know the computer could process all the information more accurately. Give more calculated options."	Subordinate
	"It gives you plenty of options but seems like it neglects some that seem more realistic, like the scenario where it always left a lane open, which is not the best thing to do."	Subordinate on highest ranked team
	"The layout of this program was very nice."	Subordinate
	"DSS seemed like a good tool."	Team leader
	"Very easy to read but boring to look at."	Team leader
	"Given as a learning tool, it started conversation and brainstorming. Gave good ideas, but not great solutions."	Team leader on team that didn't chose any computer ECOAs
	"In scenario the computer solutions worried me because it completely left one AA open. This made me lose trust in the program."	Subordinate
	"The computer may be more accurate and come up with more solutions when time is a factor."	Team leader
	"DSS helps to gather tons of info in short time, but to make the best possible decision I would not use it alone. Need human's assistance."	Subordinate
	"A useful system that can be used as an advisor to courses of action, but should not be used as a final decision maker."	Team leader
	"It really helps to look at what the computer came up with, and allowed us to make our models better. I think it is *very* useful."	Subordinate on higher performing team
	"I was much more confident with our options when the computer generated them also. It also helped visualize some options we didn't come up with."	Team leader
	"I think in modern warfare it's not as useful because our enemies use guerilla tactics, where there really are no "constraints" or courses of action set in stone."	Subordinate
Team Interaction Comments	"Team problem solving has a few downfalls, where a computer doesn't. A team that completely trusts one another will overall make the more effective decision."	Subordinate
	"It was great working with a team. I think together we were able to eliminate the obvious poor choices and talk our way to an effective method of action."	Subordinate
	"It helped to hear ourselves talk out loud."	Subordinate
	"I liked when we asked each other what did we think about it."	Subordinate
	"Because I have less experience with this and being a leader, I didn't play as much of a role in the final decision."	Subordinate from low performing team
	"I feel when we all worked together the problem got solved faster and more effectively."	Low performing team
	"It worked better when only one teammate was explaining at a time. Once all of us had something to say, it was chaotic."	Highest ranked team

Team Interaction Comments	"We all agreed on the final solutions, discussions were very effective."	Team leader
	"We talked through it to make sure we were on the same page. When drawing a solution, we all contributed. Reaching a decision, everyone revised and questioned the plan until all were satisfied."	Subordinate
	"One member did not participate fully. When attempting to engage him we wasted time with little to no response."	Team leader
	"We took the best ideas of everyone to come up with the best solutions."	Subordinate
	"I like being involved in final decision making, and had a hard time agreeing with something I didn't think was best."	Subordinate on higher performing team
Decision Making Comments	"I think the integration of computer-simulated decision making combined with human decision making provides a good choice in strategy. Eliminates unnecessary and improbable strategies and tactical plans."	Subordinate
	"Although computers do generate a very good degree of probable actions taken, I feel human can create an undetermined variable in actual warfare."	Subordinate
	"Although computers do provide a logical use in predictions in movement of troops, we cannot fully rely on it."	Subordinate
	"Decisions were better when everyone gave his or her point of view."	Subordinate

Figure 37. Participant Comments

6.12 – Summary of the Results

The following are main results from this research:

- A significant number of teams chose Weasel's flawed solution set when the system exhibited brittle behavior. This confirms a trend found in the 2004 study where individuals showed a propensity to choose the flawed set as well. This is a key result when considering how and when to utilize a DSS in domains where the consequences of the decisions are as important as those in military operations.
- On average, ECOA quality declined when Weasel exhibited brittle behavior, but there were no statistically significant differences. There seems to be a polarizing

effect of Weasel use where the results are either very good (when the system does *not* exhibit brittleness) or very bad (when the system demonstrates brittleness). Once again, this result shows the need for greater understanding of the impacts of brittle DSS behavior.

- There was a significant difference in the types of decisions made by teams versus individuals. Teams chose the automated ECOA set significantly more and individual users chose their own manually generated ECOAs more frequently.
- Weasel did not help *teams* to produce higher quality ECOAs. This result differs from that found in 2004 for *individuals* where Weasel helped users produce higher quality ECOAs. This result is also somewhat surprising based on the subjective comments from participants regarding the usefulness of the tool as well as from the researchers observations where the DSS aided teams in error checking and overall "quality control" behaviors to ensure proper ECOAs were generated.
- There was no significant difference between team and individual performance in terms of solution quality scores or rankings.
- For individuals and teams, presentation order (i.e. problem solving method) did not lead to increased preference toward computer solutions and did not impact solution quality.
- When comparing individual and team participant responses, the following significant results were found:
  - Individuals had a more favorable attitude regarding computers than did teams

- On average, the individual participants had played wargaming simulations more than the people participating in teams
- Teams were more confident than individuals that the computer would identify important ECOAs

➢ Comparison of team leaders versus subordinates demonstrated the following significant results

- Subordinates rated the quality of the DSS intelligence higher than did team leaders
- Subordinates rated their confidence in the DSS tool higher than leaders did
- Team leaders preferred to make the final decision significantly more than subordinates wanted the leader to make the decision
- Subordinates were more confident of the team effectiveness in making quality decisions than team leaders were
- Subordinates rated the overall team effectiveness higher than leaders did

➢ Higher performing teams had significantly more communication than lower performing teams

➢ A significant positive relationship existed between the overall number of ECOAs a team generated and overall solution quality scores as well as for the number of ECOAs and associated quality scores in the brittle scenario. The more ECOAs generated, the higher the solution quality score.

6.13 – Limitations

The following is a list of limitations of the research design and differences of the research design between the study involving military personnel completing the study as individuals in 2004 versus the current study (completed in 2009) where military personnel completed the study as teams.

- The research trials were completed in different buildings (UMN industrial engineering building room L121 in 2004 vs. the armory building classroom in 2009).
- The participants in 2004 were not videotaped, whereas the participants in 2009 were videotaped (and aware of the camera in the room).
- Environmental factors such as lighting, seating type, room temperature, outside noise levels, and time of day were not controlled for. Of these, some factors were very similar between the two research locations (fluorescent overhead lighting in both rooms, basic UMN table and chairs used in both rooms, and participants completed the study during the day during normal duty hours).
- Individual participant factors such as diet, sleep, exercise, and other activities of participants during the day they completed their research trial were not controlled for.

Chapter 7 – Future Work

This experiment focused on team behavior and performance when problem solving with the assistance of a DSS. Based on the results and lessons learned from this study as well as issues not analyzed (such as the impact of time constraints, team size and team member expertise), there is considerable work to be completed in future research endeavors.

Most importantly, additional research analyzing brittle DSS behavior and the impact on users should be conducted. The high number of teams who chose the flawed solution set is a very significant result of this study. In domains as life-threatening as military operations, the impact of brittle behavior must be thoroughly understood to not only facilitate the best decision making possible, but also to ensure lives are protected to the best extent possible. Considering the critical situations in which Weasel and other military DSS's are used, it is vital to know the impact brittleness may have on users.

Also, due to the apparent polarizing impact on performance and decision making behavior seen in this study, future work should focus on enhancing knowledge regarding why and when a DSS leads to very good or very poor performance. On a related issue, greater understanding is needed pertaining to why teams may choose computer generated solutions and why individual users choose their own manually generated solutions more often. Understanding the 'why's' behind such behavior may foster optimizing the use of DSS tools in many different domains.

Time constraints were not a consideration in this experiment as teams were not strictly limited in their time allotted to analyze the problem scenario and come to a decision. The time teams took to complete the scenarios was tracked, and the typical time utilized to reach a decision was approximately 20 minutes. Of the 24 scenarios completed by the twelve teams, only two were completed under 10 minutes (8:47 and 8:50), 8 were completed between 10-20 minutes, and 14 were completed between 20-30 minutes. Implementing more stringent time constraints than those used in this experiment may provide greater understanding of user decision making in a military domain. The quality of ECOAs generated and most importantly the ultimate decisions users make may be greatly influenced by the time available to subjects. In military operations time is often a limited resource and research analyzing the impact of time pressure on user decision making would be highly beneficial.

Future research may also analyze how differences in team size and level of expertise among team members impacts behavior and performance. Military personnel most often work in teams but those teams can vary greatly in size. Potential impacts exist based on team dynamics, social interaction of team members, roles of team members, and many other factors. Related to the research completed in 2004, additional analysis may also be done regarding the level of expertise among team members, as results in 2004 found Weasel helped novice users significantly more than expert users. Researching team behavior and performance with various combinations of expert and novice team members may impact the results and applicability of DSS use.

Lastly, the other components of the intel tool kit DSS (FOX, the friendly course of action generator and CoRaven, the intelligence analysis tool) have yet to be researched in detail and future research should study their utility for decision makers in the military domain. There is valuable information to be gained by analyzing FOX (the friendly course of action generator) separately as well as the with all the system components being used together. The objectives of a military force differ whether they are dealing with enemy or friendly forces. Balancing the two COA generator tools (Weasel and FOX) simultaneously adds increased complexity to an already complex situation but also provides potentially useful capabilities to the decision maker(s).

Chapter 8 – Conclusions and Recommendations

Based on analysis results and participant feedback there is great knowledge to be gained from this experiment and potential usefulness in the automated tool used. The Weasel ECOA generator is a useful aid when it generates quality solutions, and the DSS can potentially benefit military personnel with its capabilities. However, there does indeed to be a significant negative impact on user decision making when Weasel exhibited brittleness.

This leads to the most important result that decision making behavior can be negatively impacted when the DSS exhibits brittle behavior. When brittle DSS behavior leads to dangerous courses of action being selected, the results could be catastrophic. We learned from this study that teams and individual users do not necessary detect brittle behavior. Although all participants seemed to notice the fact that one AA was always left open in the Weasel COAs, in the end they did not necessarily interpret that as a dangerous or brittle recommendation (despite often initially thinking it was potentially dangerous). A primary lesson learned from this study is that although brittleness may not lead to lower quality ECOAs from the users, the quality of the eventual decision may absolutely be of lower quality when brittleness is present. These results highlight the importance of further research investigating the influence of brittle DSS behavior on the system's user(s).

It is encouraging to see that Weasel can positively impact ECOA solution quality when the DSS generates high quality (not brittle) solutions. Used as an "idea" or solution generator and quality control mechanism, objective and subjective data show users find a definite benefit in Weasel. An interesting consideration however is the apparent polarizing effect using Weasel can have. When the system is not brittle, it clearly helps, but when it is brittle in generating solutions, there is a definite negative impact. This seems intuitive but is none the less very important to understand since developers and users typically don't know *when* a DSS will demonstrate brittleness. Implications of this polarizing effect are that design of DSSs must attempt to minimize brittleness and training should facilitate user's recognition of and avoidance in adopting potentially brittle DSS solutions.

Another very useful result pertaining to team DSS use demonstrated that higher performing teams had significantly more communication than lower performing teams. This result coincides with those found in the Jin and Levis study discussed in the literature review, where a hierarchical organization structure (like that used in the military), led to improved team interaction and communication [7]. In this study, good communication seemed to foster understanding the problem, developing valid, higher quality courses of actions, and overall team decision making. This result is very useful when analyzing how to implement DSS use with military teams. Such a system should be utilized in situations where team interaction and communication is possible and promoted to the highest degree available.

Also, a significant positive relationship was found to exist between the number of ECOAs a team generated and solution quality scores. The more ECOAs generated, the higher the solution quality score. Again, this result helps researchers and DSS implementers understand what types of problem solving processes may be best when users are working with a DSS. The fact of simply generated an adequately large set of courses of action may increase the eventual quality of the decision and course of action chosen. User may be made aware of this fact and the idea of brainstorming and thinking of different courses of action, regardless of initial consideration of quality or validity, may lead to improved decision making.

Results pertaining to team versus individual performance showed there was no difference between teams and individuals. This may be due to teams being better problem solvers or COA generators to begin with, therefore the DSS does not help them as much as individuals. Individuals may not perform as well without DSS assistance but they do perform better (and just as well as teams) with the use of Weasel. Based on the result that individuals performed better with Weasel, along with the subjective feedback from users and observations, there seems to be a definitive benefit for individuals as well as teams to utilize DSS tools. If only to generate potential COAs or facilitate ideas and analysis, the benefits of a DSS are apparent. There are multiple benefits to DSS use. Users could save time if the system can recommend COAs to use as a starting point for additional COA generation and analysis. A DSS can interpret much greater amounts of intelligence and other data than a human can, resulting in potentially better solutions in highly complex problems. As seen in this study, the DSS can act as a sort of quality control mechanism for users to ensure constraints are being abided by. Of course, as the results pertaining to DSS brittleness highlight, there are pit falls to be aware of (and hopefully addressed via effective DSS design and implementation). Users may over-rely on the automation for adequate solutions to complex problems. Other alternative and potentially better solutions may not be considered if users focus to heavily on COAs generated by the DSS. Most importantly, the consequences of choosing poor quality DSS solutions can lead to disastrous results in the field.

Additional useful insight was gained regarding user attitudes, team interaction and their relationship to decision making behavior. Many of these results show interesting social dynamics between team leaders and subordinates. Participants in the leader role appear to like being in charge and having the responsibility of making decisions, while subordinates value being involved in the process as well and prefer that the entire team make the final decision (showing a methodology of decision by consensus may be preferred over a dictatorship approach).

Given the results of this study, along with those found in the 2004 research, the DSS has a definite utility as a training tool for military personnel (for both individuals and teams) in battlefield decision making and strategies. This is due to the objective data as well as the subjective input of all participants that Weasel provided great assistance in generating military strategies and was a valuable tool to practice decision making in military operation scenarios. It is quite amazing to see military personnel, with little to no previous experience in tasks such as those implemented in this study, being able to strategize, develop courses of action, and work towards making a high quality decision within a matter of minutes when working with Weasel. In my view, the potential benefit of personnel using Weasel and other DSS applications is tremendous. However, training and implementation of such systems must clearly focus on recognition of brittle behavior and all attempts to minimize the impacts of that brittleness should be taken.

Further study is recommended pertaining to Weasel being utilized concurrently with the other automation tools that accompany the Intel Tool Kit (specifically FOX, the friendly course of action tool). Additionally, further analysis on the impact of brittleness and apparent performance extremes are two areas where additional research should focus.

This study, along with the work completed in 2004, has shed significant light on the applicability and usefulness of Weasel on military decision making. However, there is much more research to be conducted to fully understand how humans make decisions and what impacts their performance and behavior. This research provided important results to enhance academic and operational understanding of decision support systems. The desire is that the results and recommendations of this study provide further impetus towards DSS development, research, and implementation to enhance human decision making.

Bibliography

[1] Bisantz, Ann M. and Younho Seong, 2001. "Assessment of Operator Trust in and Utilization of Automated Decision-Aids Under Different Framing Conditions." *International Journal of Industrial Ergonomics*. Vol. 28: 85-97.

[2] Cohen, M. S., J.T. Freeman, and S. Wolf, 1996. "Metarecognition in Time-Stressed Decision Making: Recognizing, Critiquing, and Correcting." *Human Factors.* Vol. 38, No. 2: 206-219.

[3] Cook, Malcolm, L. Elder, and G. Ward, 1997. "Decision Making, Planning, and Teams." *Institution of Electrical Engineers*. Vol. 5, pp. 1-22.

[4] Cowan, T. A., 1963. "Decision Theory in Law, Science, and Technology." *Science*. Vol. 140, No. 3571: 1065-1075.

[5] Gallupe, R. Brent and Gerardine DeSanctis and Gary Dickson, 1988. "Computer-based support for group problem-finding: an experimental investigation." *MIS Quarterly*. Vol. 12, No. 2, pp. 277-296.

[6] Holloman, Charles and Hal Hendrick, 1972. "Adequacy of group decisions as a function of the decision-making process." *The Academy of Management Journal*. Vol. 15, No. 2, pp. 175-184.

[7] Jin, Victoria and Alexander Levis, 1990. "Compensatory behavior in team decision making." *IEEE*. pp. 107-112.

[8] Jones, Brian S., 1997. "Co-operative decision making and team work in command and control: how can computers help and how do they get in the way?" *Institution of Electrical Engineers*. Vol. 6, pp. 1-5.

[9] Larson, Adam D. and Caroline Hayes, 2005. "An Assessment of Weasel: A Decision Support System to Assist in Military Planning" *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*. pp. 287-291.

[10] Lehner, Paul et al, 1997. "Cognitive biases and time stress in team decision making." *IEEE Transactions on Systems, Man, and Cybernetics*. Vol. 27, No. 5, pp. 698-703.

[11] Lee, John D. and Neville Moray, 1992. "Trust, Control Strategies, and Allocation of Function in Human-Machine Systems." *Ergonomics*. Vol. 35, pp. 1243-1270.

[12] Lee, J.D. and Moray, N., 1994. "Trust, Self-confidence, and Operators' Adaptation to Automation." *International Journal of Human-Computer Studies*. Vol. 40, pp. 153-184.

[13] McClumpha, A. and M. James, 1994. "Understanding Automated Aircraft." *Human Performance in Automated Systems: Recent Research and Trends*. pp. 314-319.

[14] Ravinder, U., 2003. "Weasel: A constraint-based tool for generating Enemy Courses of Action." Master's Thesis.

[15] Parasuraman, Raja and Victor Riley, 1997. "Humans and Automation: Use, Misuse, Disuse, Abuse." *Human Factors*. Vol. 39, No. 2, pp. 230-253.

[16] Riley, Victor, 1989. "A General Model of Mixed-Initiative Human-Machine Systems." *Proceedings of the Human Factors Society 33rd Annual Meeting.* pp. 124-128.

[17] Schlabach, J.L., C.C. Hayes, and D.E. Goldberg, 1998. “FOX-GA: A Genetic Algorithm for Generating and Analyzing Battlefield Courses of Action.” *Evolutionary Computation*. Vol. 7, No. 1, pp. 45-68.

[18] Sheridan, Thomas and William Verplank, 1978. “Human and Computer Control of Undersea Teleoperators.” Cambridge, MA. Massachusetts Institute of Technology.

[19] Shim, J.P. et al., 2002. “Past, present, and future of decision support technology” *Decision Support Systems*. Vol. 33, pp. 111-126.

[20] Smith, Philip J., C. Elaine McCoy, and Charles Layton, 1997. “Brittleness in the Design of Cooperative Problem-Solving Systems: The Effects on User Performance.” *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans.* Vol. 27, No. 3, pp. 360-371.

[21] Steeb, R. and S.C. Johnston, 1981. “A computer-based interactive system for group decision making.” *IEEE Transactions on Systems, Man, and Cybernetics.* pp. 544-552.

[22] Swets, John A, 1973. “The Relative Operating Characteristic in Psychology.” *Science*. Vol. 182, No. 4116, pp. 990-1000.

[23] Thorsden, Marvin and Gary Klein, 1990. “Cognitive features of team decision making” *IEEE*, pp. 891-893.

[24] Tjosvold, Dean and Richard Field, 1983. “Effects of social context on consensus and majority vote decision making.” *The Academy of Management Journal*. Vol. 26, No. 3, pp. 500-506.

[25] Trull, Samuel, 1966. "Some factors involved in determining total decision success" *Management Science*. Vol. 12, No. 6, pp. B270-B280.

[26] Turoff, M. and S.R. Hiltz, 1982. "Computer support for group versus individual decisions." *IEEE Transactions on Communications*. pp. 82-90.

[27] Weisstein, Eric W. "Spearman Rank Correlation Coefficient." From Mathworld - A Wolfram Web Resource. http://mathworld.wolfram.com/SpearmanRankCorrelationCoefficient.html

[28] Wiegmann, Douglas A., 2002. "Agreeing With Automated Diagnostic Aids: A Study of Users' Concurrence Strategies." *Human Factors and Ergonomics Society*. Vol. 44, No. 1, pp. 44-50.

[29] Wierenga, Berend and Peter Ophuis, 1997. "Marketing decision support systems: Adoption, use, and satisfaction." *International Journal of Research in Marketing*. Vol. 14, pp. 275-290.

[30] www.fas.org/man/dod-101/army/docs/fm101-5-1/f545-c4a.htm

Appendices

Appendix 1

Appendix 1.1 IRB Approval Notification

Appendix 1.1 displays the approval letter from the University of Minnesota institutional review board (IRB).

Subject: **0904E64047 - PI Larson - IRB - Exempt Study Notification**

From: irb@umn.edu

Date: Tue, 5 May 2009 16:12:00 -0500 (CDT)

To: lars1770@umn.edu

TO : lars1770@umn.edu,

The IRB: Human Subjects Committee determined that the referenced study is exempt from review under federal guidelines 45 CFR Part 46.101(b) category #2 SURVEYS/INTERVIEWS; STANDARDIZED EDUCATIONAL TESTS; OBSERVATION OF PUBLIC BEHAVIOR.

Study Number: 0904E64047

Principal Investigator: Adam Larson

Title(s):
The Impact of Computer Decision Support on Military Team Decision Making

---

This e-mail confirmation is your official University of Minnesota RSPP notification of exemption from full committee review. You will not receive a hard copy or letter.
This secure electronic notification between password protected authentications has been deemed by the University of Minnesota to constitute a legal signature.

The study number above is assigned to your research. That number and the title of your study must be used in all communication with the IRB office.

Research that involves observation can be approved under this category without obtaining consent.

SURVEY OR INTERVIEW RESEARCH APPROVED AS EXEMPT UNDER THIS CATEGORY IS LIMITED TO ADULT SUBJECTS.

This exemption is valid for five years from the date of this correspondence and will be filed inactive at that time. You will receive a notification prior to inactivation. If this research will extend beyond five years, you must submit a new application to the IRB before the study?s expiration date.

Upon receipt of this email, you may begin your research. If you have questions, please call the IRB office at (612) 626-5654.

You may go to the View Completed section of eResearch Central at http://eresearch.umn.edu/ to view further details on your study.

The IRB wishes you success with this research.

https://gophermail.umn.edu/session/lars1770//AAAK@display@32@41?print=friendly Page 1 of 1

Appendix 1.2 Participant Consent Form

Appendix 1.2 contains the consent form presented to participants immediately upon their arrival to the study. Each participant read and signed two copies of the consent form prior to beginning the experiment. One copy of the consent form was provided to each participant.

CONSENT FORM
for
The Impact of Computer Decision Support on Military Team Decision Making

You are invited to be in a research study of decision making and automation use in a military battlefield simulation. You were selected as a possible participant because you are a member of the military with the necessary background to adequately participate in the simulation. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by: Capt Adam Larson, Univ. of Minnesota Department of Industrial and Systems Engineering

Background Information

The purpose of this study is to gain a better understanding of how people make decisions when weighing computer-generated options versus their own and when decision making is limited by time. This research will provide useful information not only in overall decision making processes, but also to evaluate military decision making methods and how time constraints affect decisions.

Procedures:

If you agree to be in this study, we would ask you to do the following things:
Provide approximately 2 total hours of your time to be trained in the simulation program and perform the necessary experimental trials. You will sit at a table and perform the simulation based on the scenarios provided by the researcher and the information presented in the computer simulation. You will also be asked to complete a 5-10 minute survey regarding confidence levels and trust in automation.

Risks and Benefits of being in the Study

The study has no physical risks and no direct benefits.

Compensation:
None

Confidentiality:

The records of this study will be kept private. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota or with the ROTC detachment. If you decide to participate, you are free to not answer any question or withdraw at any time with out affecting those relationships.

Contacts and Questions:

The researcher conducting this study is: Adam Larson, graduate student, U of MN/Industrial Engineering. You may ask any questions you have now. If you have questions later, **you are encouraged** to contact them at Room L121, 111 Church Street, Minneapolis MN 55455, or via phone at 651-459-0302, or via email at lars1770@umn.edu.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

You will be given a copy of this information to keep for your records.

Statement of Consent:

I have read the above information. I have asked questions and have received answers. I consent to participate in the study.

Signature: ____Date: __________________

Signature of Investigator: ____________________________________Date: __________________

Appendix 2

Appendix 2.1 Scenario 1 Description and Weasel ECOAs

Appendix 2.1 is the situational description given to participants for scenario 1 followed by the 8 computer generated ECOAs from Weasel.

<u>Scenario 1</u>

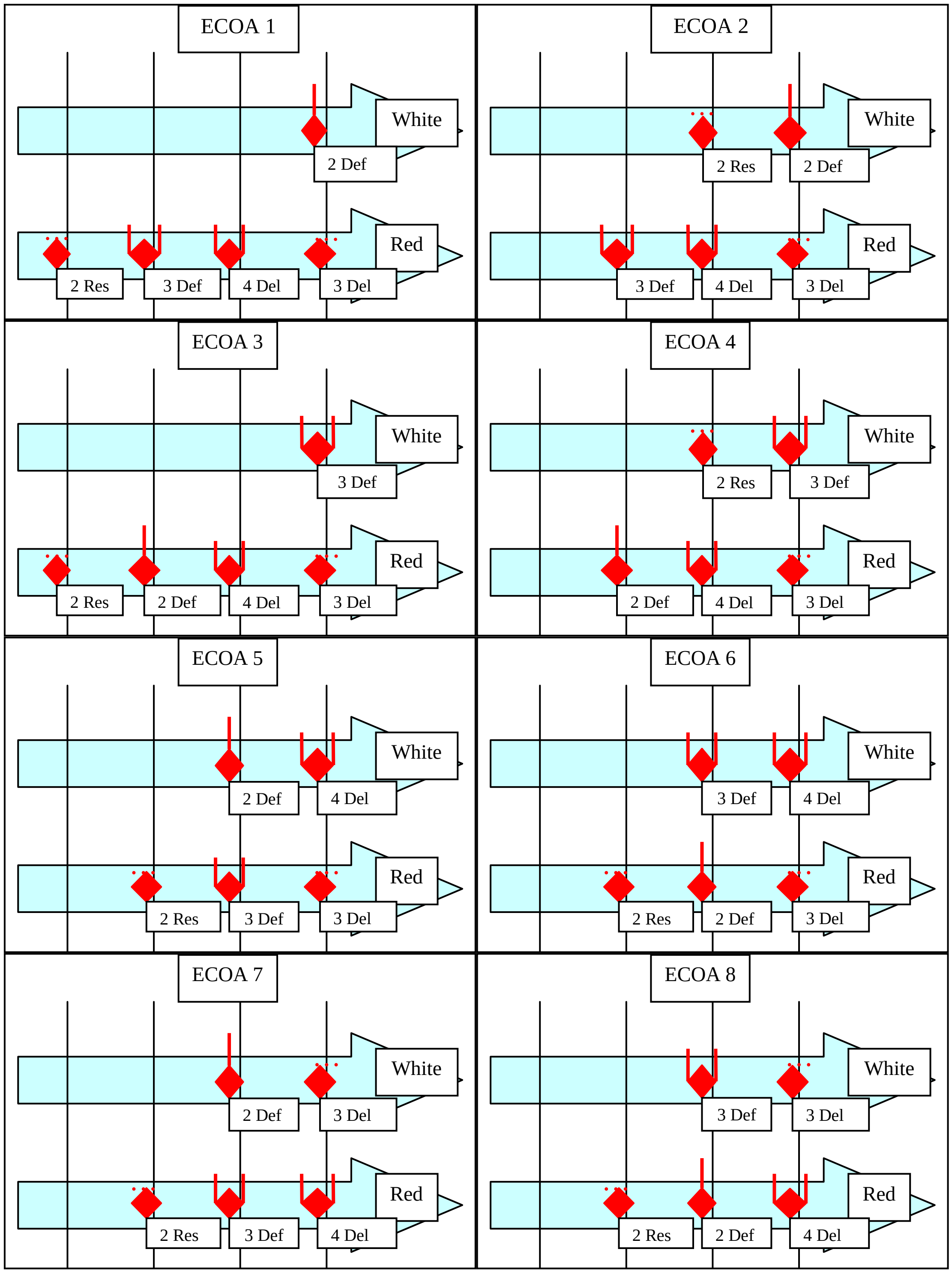
You've received intelligence from allied ground troops that enemy forces are orchestrating a massive defense in anticipation of being attacked by your allied forces. You want to identify possible enemy defenses so friendly forces know what they may encounter. You know the following about the enemy's situation:

- ❖ Intention is for forces to defend 2 Avenues of Approach (Axis White is to the north, Axis Red is the southernmost AA).

- ❖ There are a large number of enemy forces, made up of 2 battalions, 2 platoons, and 1 company. Details regarding each unit follow:
  - 1 battalion is committed to defend with 3 motorized subunits.
  - 1 company is committed to defend with 2 armored subunits.
  - 1 battalion is defending in delay with 4 mechanized infantry subunits.
  - 1 platoon is defending in delay with 3 armored subunits.
  - 1 platoon is defending in reserve with 2 motorized subunits.

- ❖ The main effort of defense will be on Axis Red, the southern avenue of approach.

- ❖ Forces can be as deep as line of defensible terrain (LDT) 2.


ECOA 1
White
2 Def
Red
2 Res
3 Def
4 Del
3 Del
ECOA 2
White
2 Res
2 Def
Red
3 Def
4 Del
3 Del
ECOA 3
White
3 Def
Red
2 Res
2 Def
4 Del
3 Del
ECOA 4
White
2 Res
3 Def
Red
2 Def
4 Del
3 Del
ECOA 5
White
2 Def
4 Del
Red
2 Res
3 Def
3 Del
ECOA 6
White
3 Def
4 Del
Red
2 Res
2 Def
3 Del
ECOA 7
White
2 Def
3 Del
Red
2 Res
3 Def
4 Del
ECOA 8
White
3 Def
3 Del
Red
2 Res
2 Def
4 Del

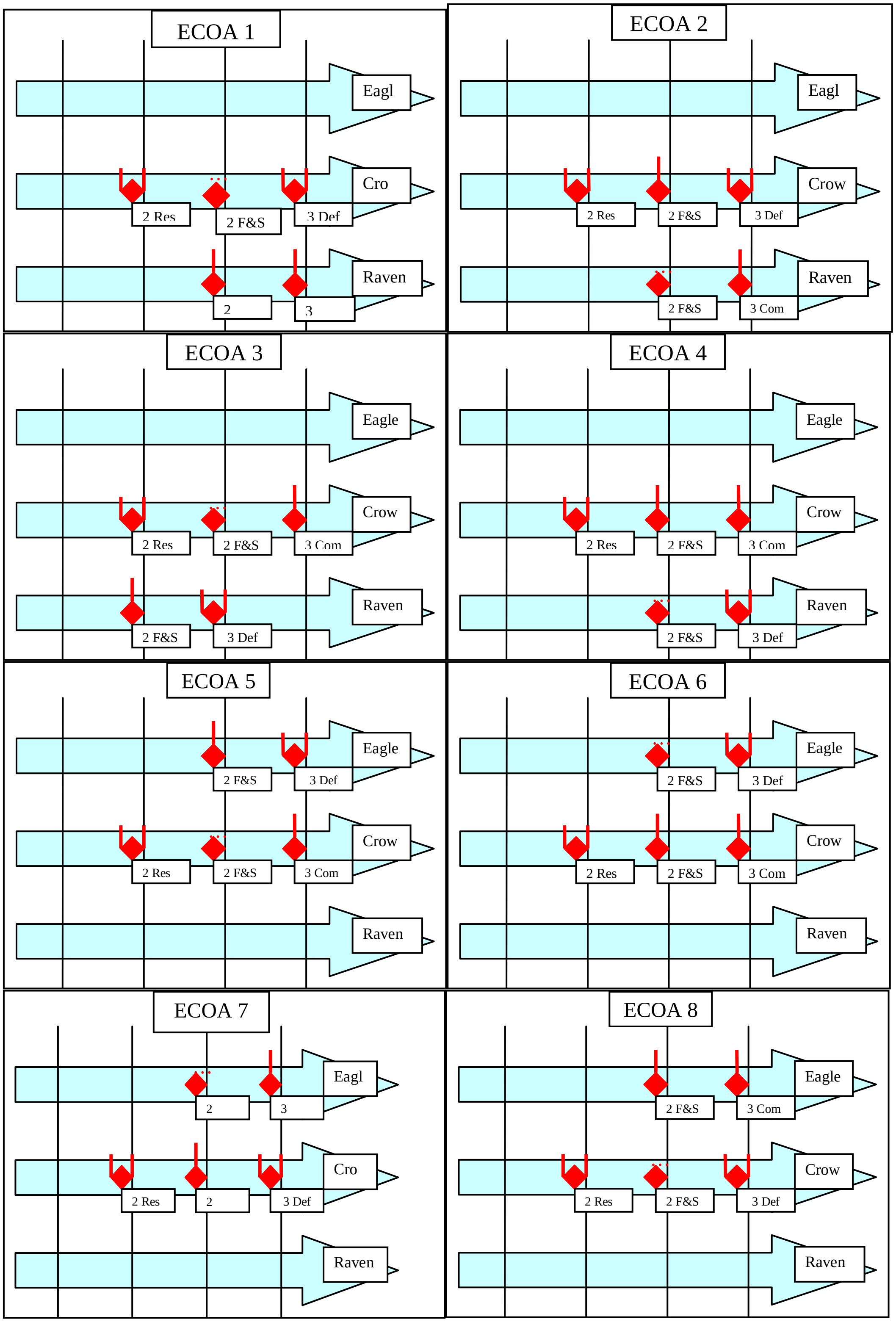
Appendix 2.2 Scenario 2 Description and Weasel ECOAs

Appendix 2.2 is the situational description given to participants for scenario 2 followed by the 8 computer generated ECOAs from Weasel.

Scenario 2

You've received intelligence from allied ground troops indicating enemy action of some kind. The enemy has been mobilizing troops over the last 48 hours and seems to be capable of taking various actions. Enemy intelligence shows:

- ❖ Forces are in place to attack and defend along 3 Avenues of Approach (Eagle is northernmost AA, Crow is in the middle, and Raven is southernmost AA).
- ❖ Enemy forces consist of 2 companies, 2 battalions, and 1 platoon. Details on each unit follows:
  - ➢ Attacking forces
    - 1 company is committed to attack [C] with 3 armored subunits
    - 1 company is following and supporting (F&S) in attack mode with 2 motorized subunits.
    - 1 platoon is also following and supporting (F&S) in attack mode with 2 mechanized infantry subunits
    - ▪ Note: Intercepting communication lines tells you the enemy will not place more than 1 F&S unit on the same AA.
  - ➢ Defense forces
    - 1 battalion is committed to defend (Def) with 3 mechanized infantry subunits
    - 1 battalion is defending in reserve (R) with 2 motorized subunits
- ❖ The enemy plans on concentrating their main effort on AA Crow (middle).
- ❖ Forces can be as deep as line of defensible terrain (LDT) 3.


ECOA 1
Eagl
Cro
2 Res
2 F&S
3 Def
Raven
2
3
ECOA 2
Eagl
Crow
2 Res
2 F&S
3 Def
Raven
2 F&S
3 Com
ECOA 3
Eagle
Crow
2 Res
2 F&S
3 Com
Raven
2 F&S
3 Def
ECOA 4
Eagle
Crow
2 Res
2 F&S
3 Com
Raven
2 F&S
3 Def
ECOA 5
Eagle
2 F&S
3 Def
Crow
2 Res
2 F&S
3 Com
Raven
ECOA 6
Eagle
2 F&S
3 Def
Crow
2 Res
2 F&S
3 Com
Raven
ECOA 7
Eagl
2
3
Cro
2 Res
2
3 Def
Raven
ECOA 8
Eagle
2 F&S
3 Com
Crow
2 Res
2 F&S
3 Def
Raven

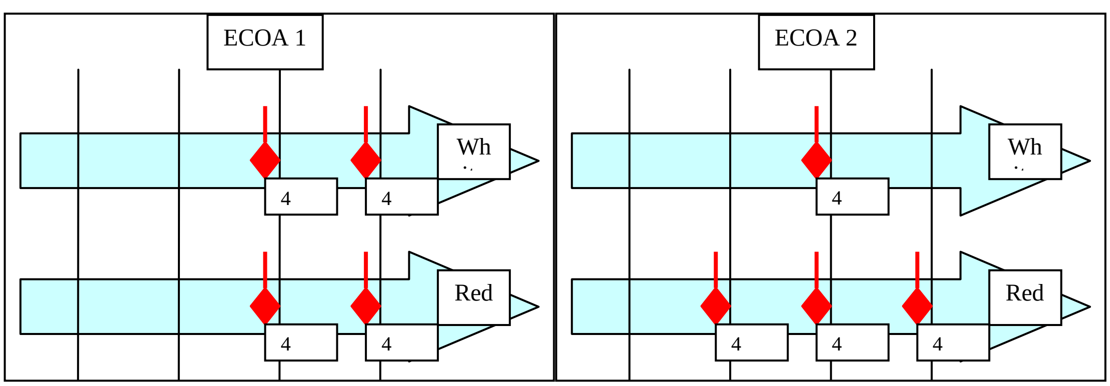
Appendix 2.3 Scenario 3 Description and Weasel ECOAs

Appendix 2.3 is the situational description given to participants for scenario 3, which was the practice scenario, followed by the 2 computer generated ECOAs from Weasel.

<u>Scenario 3</u>

You've received intelligence from computer surveillance and analysis that enemy forces are on the move, heading towards your allied forces. To best prepare your defense, you must identify any possible courses of action the enemy may take in attacking your forces. You are confident in the following enemy intelligence:

- ❖ Intention is to attack from the west along 2 Avenues of Approach (AA_white is north of AA_red).
- ❖ Enemy forces consist of 4 companies, each with 4 subunits.
    - 2 companies are committed to attack [C] with armored subunits
    - 2 companies might attack in reserve (R) with motorized subunits.
- ❖ The enemy plans on concentrating their main effort on AA_red.
- ❖ Forces can be as deep as LDT 3.



Appendix 2.4 Questionnaires

Appendix 2.4 contains the questionnaires given to each participant after the scenarios were completed.

Questionnaire for

"Simulation Decision Making and Trust in Automation"

Name: ________________________ **Date: __________**

Please **circle** your most accurate response to the following items:

1. **Rate your general level of trust in computers:**

1	2	3	4	5
Low Trust		Average Trust		High Trust

2. **Rate your trust in computer analysis to determine military intelligence and tactics.**

1	2	3	4	5
Low Trust		Moderate Trust		High Trust

3. **How would you generally rate your trust in computer-generated solutions to a problem vs. your own personal solutions to the same problem?**

0	1	2	3	4	5
Insufficient Information to Answer	Trust Computer Solution More		Equal Trust		Trust My Personal Solution More

4. **What rating best represents your attitude about computers?**

1	2	3	4	5
Completely dislike computers		Average		Completely like computers

5. **On average, how many hours per week do you play any type of computerized wargame simulation?**

1	2	3	4	5
0 hrs	1-2 hrs	3-5 hrs	6-10 hrs	More than 10 hrs

6. **Rate your confidence in identifying all important enemy courses of action in this experiment.**

1	2	3	4	5
Low Confidence		Moderate Confidence		High Confidence

7. **Rate your confidence in the computer identifying important enemy courses of action in this experiment.**

1	2	3	4	5
Low Confidence		Moderate Confidence		High Confidence

8. **Rate your confidence in the decisions you made when choosing between automated and personal enemy courses of action for this experiment.**

1	2	3	4	5
Low Confidence		Moderate Confidence		High Confidence

9. **In your opinion, who would make a better / more trustworthy decision in a situation with multiple variables and potential risk involved?**

Human Computer

Survey for

“DSS Analysis”

- ROTC Year (circle): Freshman Sophomore Junior Senior

- Any previous experience working with a DSS (circle): Yes No

- Team Role (circle): Leader Subordinate

- Date:__________

Place an “X” in one of the following categories as it relates to working with **the specific DSS tool you just used.**					
	Very Poor 1	**Poor 2**	**Average 3**	**Good 4**	**Very Good 5**
Quality of displays					
Quality of intelligence info					
Quality of courses of action					
Quality of information organization					
Quality of appearance or style					
Ease of use					
Quality as a decision making tool					
Level of innovation					
Confidence in the tool					
Overall usefulness					
Overall quality					

Please add any comments regarding the **design, usefulness, quality, etc. of the specific DSS** used in this study:

Place an “X” in one of the following categories as it relates to working with **your team. (note: different rating scale)**					
	Strongly Disagree 1	**Somewhat Disagree 2**	**Neutral 3**	**Somewhat Agree 4**	**Strongly Agree 5**
I trusted the decision making input of my other team members					
Subordinates only: I trusted the decision making of the team leader					
I preferred to be involved in the final decision					
I preferred the team leader make the final decision					
Communication was effective when everyone made the decision					
Communication was effective when the team leader made the decision					
A better final decision resulted when all team members made the decision					
A better final decision resulted when the team leader made the decision					
Overall, our team was effective in the quality of our decisions					
Overall, our team was effective in communicating with one another					
Overall, our team was effective in working together					

Please add any comments and recommendations regarding the **effectiveness of decision making, communication, team problem solving, etc.** based on your experience in this study: ____

____

____

____

Appendix 2.5 Data Recording Form

This appendix contains the form used by the researcher to record participant data as each experiment trial was conducted.

Name / Team #:			**Subject #:**	
Date:				
Military Service:	Air Force			
Scenario 1:				
	# Personal ECOAs generated:			
Time:	# Personal ECOAs revised:			
	ECOA set chosen: Personal	Automated		Both
	Reason for decision:			
	Verbal comments during scenario 1:			
Scenario 2:				
	# Personal ECOAs generated:			
Time:	# Personal ECOAs revised:			
	ECOA set chosen: Personal	Automated		Both
	Reason for decision:			
	Verbal comments during scenario 2:			
Comments / Observations:				

Appendix 2.6 Explanation of Study

Appendix 2.6 shows the brief explanation of the research study being conducted. This page was read by each participant prior to beginning familiarization training on the DSS tool.

<u>Simulation Decision Making and Trust in Automation</u>

Research by: Capt Adam Larson, USAF

1. The purpose of this research is to gain a better understanding of how people make decisions when weighing computer-generated options versus their own. The research will provide useful information not only in overall decision making processes, but also to evaluate and gain insight regarding military decision making methods.

2. The research uses an Army battlefield simulation that has been developed over recent years by Prof Caroline Hayes from the industrial engineering department. The Air Force and Army have helped fund this research and it will be useful in evaluating decision making for all military services.

3. The task of subjects will be to evaluate intelligence information, formulate potential enemy courses of action, and analyze courses of action generated by the computer. Considering the given circumstances outlined in the scenario and all relevant information, a decision will then be made to choose the best set of enemy courses of action. The total time for volunteering (to complete training and experiment trials) will be about 2 hours.

4. Your time in helping with this research is greatly appreciated. Please contact Capt Adam Larson with questions or comments:

Respectfully,
Adam Larson, Capt, USAF

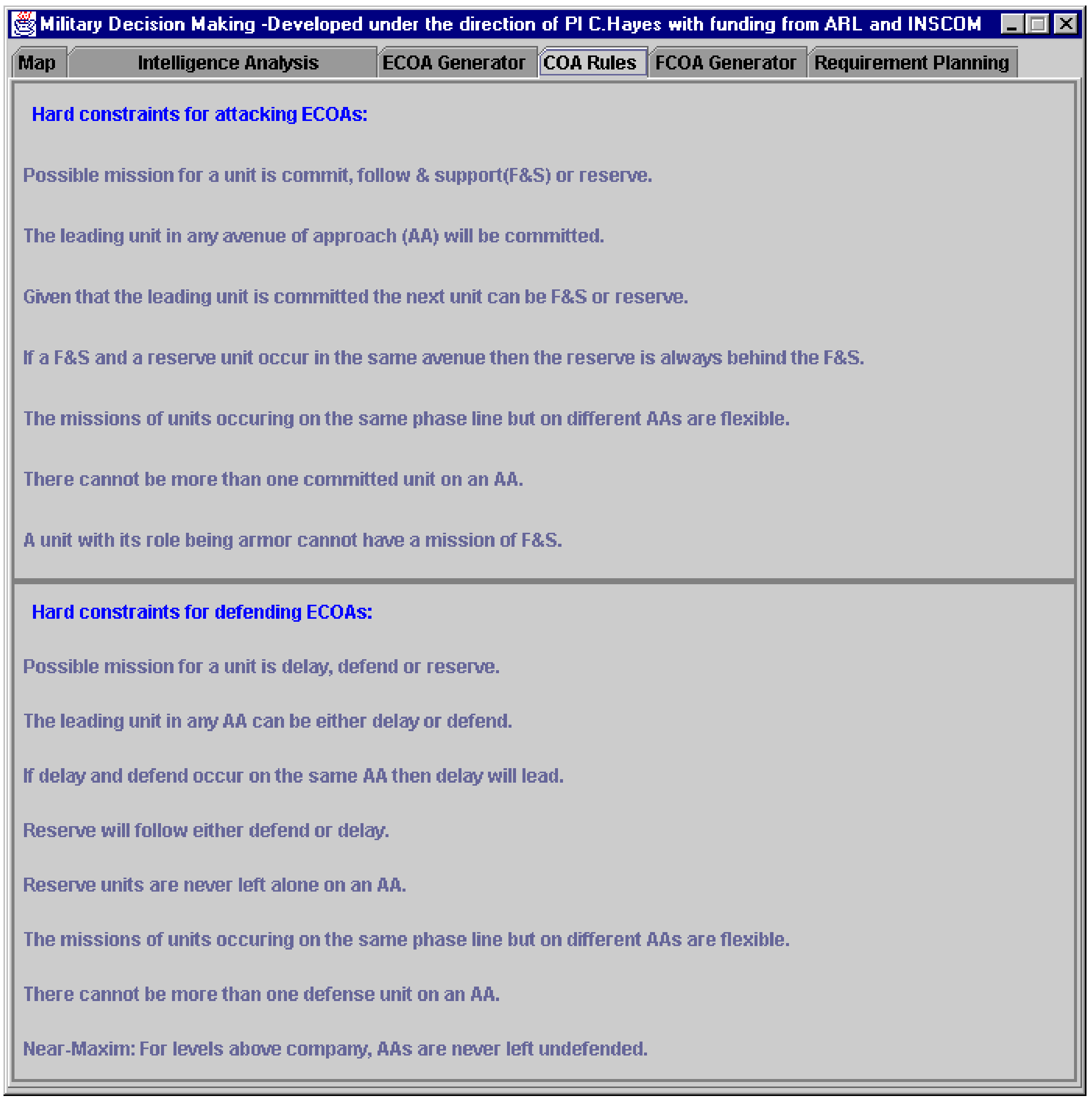
Appendix 2.7 Acronym List

The following is an acronym list of typical acronyms used in this research.

AA	Avenue of Approach
AF	Air Force
ANOVA	Analysis of Variance
C	Commit (Weasel symbology)
COA	Course of Action
DATK	Deliberate Attack
Def	Defend (Weasel symbology)
Del	Delay (Weasel symbology)
DP	Decision Point
DSS	Decision Support System
ECOA	Enemy Course of Action
EIS	Executive Information System
FCOA	Friendly Course of Action
Fox	Generates possible courses of action for friendly forces
F&S	Follow and Support (Weasel symbology)
GA	Genetic Algorithm
GDSS	Group Decision Support System
IRB	Institutional Review Board
LDT	Line of Defensible Terrain
MTC	Movement to Contact
NAI	Named Area of Interest
R	Reserve (Weasel symbology)
R/M	Recognition / Metacognition
ROTC	Reserve Officer Training Corp
RPD	Recognition-Primed Decision
Weasel	Generates potential courses of action for enemy forces

Appendix 2.8 Weasel DSS Constraints

Appendix 2.8 shows the one page screen capture that was provided to each participant. These constraints were discussed during familiarization training and participants were instructed to abide by them to the best of their abilities when generating and analyzing ECOAs. Participants were provided as much time as they needed to study and understand these constraints before beginning the scenario tasks.



Appendix 3

Appendix 3.1 Spearman Rank Correlation calculations

This appendix contains the data for the Spearman Rank Correlation value for each of the two scenarios discussed in section 6.1.

SPEARMAN RANK CORRELATION – SCENARIO 1				
Subj #	Eval 1 Rank	Eval 2 Rank	(d2)	d2/N(N2 -1)
1	7.5	11.5	16	0.000593
2	14.5	13	2.25	8.34E-05
3	18.5	21	6.25	0.000232
4	29	30	1	3.71E-05
5	9.5	4	30.25	0.001122
6	2	9.5	56.25	0.002086
7	5.5	6	0.25	9.27E-06
8	26	28.5	6.25	0.000232
9	14.5	11.5	9	0.000334
10	20	20	0	0
11	27	26	1	3.71E-05
12	7.5	3	20.25	0.000751
13	12	7.5	20.25	0.000751
14	21.5	15.5	36	0.001335
15	29	27	4	0.000148
16	21.5	22	0.25	9.27E-06
17	29	28.5	0.25	9.27E-06
18	1	5	16	0.000593
T1	17	19	4	0.000148
T2	18.5	18	0.25	9.27E-06
T3	23.5	24.5	1	3.71E-05
T4	25	24.5	0.25	9.27E-06
T5	5.5	2	12.25	0.000454
T6	16	17	1	3.71E-05
T7	23.5	23	0.25	9.27E-06
T8	3	1	4	0.000148
T9	13	15.5	6.25	0.000232
T10	4	9.5	30.25	0.001122
T11	9.5	7.5	4	0.000148
T12	11	14	9	0.000334
			Scenario 1 r_s =	**0.933**

SPEARMAN RANK CORRELATION – SCENARIO 2				
Subj #	Eval 1 Rank	Eval 2 Rank	(d2)	d2/N(N2 -1)
1	4	6	4	0.000148
2	5.5	3	6.25	0.000232
3	10	8.5	2.25	8.34E-05
4	23.5	23.5	0	0
5	13	11	4	0.000148
6	5.5	7	2.25	8.34E-05
7	29	21	64	0.002373
8	15	10	25	0.000927
9	22	23.5	2.25	8.34E-05
10	7	13	36	0.001335
11	10	8.5	2.25	8.34E-05
12	17	12	25	0.000927
13	13	16	9	0.000334
14	10	15	25	0.000927
15	27	28	1	3.71E-05
16	27	29.5	6.25	0.000232
17	30	27	9	0.000334
18	8	14	36	0.001335
T1	3	4	1	3.71E-05
T2	25	25.5	0.25	9.27E-06
T3	27	29.5	6.25	0.000232
T4	20.5	18.5	4	0.000148
T5	1	1	0	0
T6	23.5	25.5	4	0.000148
T7	13	5	64	0.002373
T8	2	2	0	0
T9	19	22	9	0.000334
T10	16	17	1	3.71E-05
T11	20.5	18.5	4	0.000148
T12	18	20	4	0.000148
			Scenario 2 r_s =	**0.921**

SPEARMAN RANK CORRELATION – OVERALL RANKINGS				
Subj #	Eval 1 Rank	Eval 2 Rank	(d2)	d2/N(N2 -1)
1	5	6	1	3.70782E-05
2	6.5	3	12.25	0.000454208
3	13	13	0	0
4	28	28.5	0.25	9.26956E-06
5	8	4	16	0.000593252
6	3	6	9	0.000333704
7	17.5	15	6.25	0.000231739
8	23	22.5	0.25	9.26956E-06
9	20	18.5	2.25	8.3426E-05
10	11	17	36	0.001334816
11	21.5	20.5	1	3.70782E-05
12	12	6	36	0.001334816
13	11	10	1	3.70782E-05
14	14.5	16	2.25	8.3426E-05
15	29	30	1	3.70782E-05
16	25.5	26	0.25	9.26956E-06
17	30	28.5	2.25	8.3426E-05
18	4	9	25	0.000926956
T1	6.5	8	2.25	8.3426E-05
T2	24	24	0	0
T3	27	27	0	0
T4	25.5	25	0.25	9.26956E-06
T5	2	1	1	3.70782E-05
T6	21.5	22.5	1	3.70782E-05
T7	19	11	64	0.002373007
T8	1	2	1	3.70782E-05
T9	17.5	20.5	9	0.000333704
T10	9	13	16	0.000593252
T11	16	13	9	0.000333704
T12	14.5	18.5	16	0.000593252
			Overall r_s =	**0.939**

Appendix 3.2 ANOVA Calculations

The tables below show the ANOVA analysis discussed throughout Chapter 6.

ANOVAs for Section 6.2

Replication of 2004 results and 2009 results for individual users

- *Does Weasel help individual users overall to produce better ECOAs?*

<u>2004 Evaluators</u>

Anova: Single Factor		**2004: Individual Quality Scores With vs. Without Weasel (Method B/C vs. Method A)**				
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Qual. Score w/out Weasel (Meth A)	18	92.25	5.125	2.71507		
Ave. Score w/ Weasel (Meth B/C)	18	112.375	6.24305	0.93285335		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	11.250	1	11.2504	6.16812	**0.0180976**	4.130017
Within Groups	62.014	34	1.8239			

<u>2009 Evaluators</u>

Anova: Single Factor		**2009: Individual Quality Scores With and Without Weasel**				
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Qual. Score w/out Weasel (Meth A)	18	91.5	5.083333	2.3970		
Ave. Score w/ Weasel (Meth B/C)	18	111	6.166667	1.4705		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	10.5625	1	10.5625	5.4619	**0.025462**	4.130018
Within Groups	65.75	34	1.933824			

- *Does ECOA quality decline for individuals when Weasel exhibits brittle behavior?*

2004 Evaluators

Anova: Single Factor — **Brittle Scenario: Individual Quality Averages With & Without Weasel**

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	6	32.75	5.458333	3.085417
Column 2	12	71.5	5.958333	1.75947

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	1	1	1	0.460018	**0.507307**	4.493998
Within Groups	34.78125	16	2.173828			
Total	35.78125	17				

2009 Evaluators

Anova: Single Factor — Quality Scores in Brittle Scenario

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Quality Without Weasel (meth A)	6	31.75	5.291667	2.860417
Quality with Weasel (B/C)	12	70.25	5.854167	1.971117

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	1.265625	1	1.265625	0.562744	**0.464046**	4.493998
Within Groups	35.98438	16	2.249023			
Total	37.25	17				

- *Does presentation order impact performance?*

2004 evaluators

2004 Evaluators: Average Individual Quality Scores for Method B vs Method C

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Method B	18	102.25	5.68055	2.851511
Method C	18	122.5	6.80555	0.849673

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	11.390625	1	11.3906	6.155123	**0.018210**	4.13001
Within Groups	62.92013	34	1.850592			
Total	74.31076	35				

2009 Evaluators

2009 Evaluators: Average Individual Quality Scores for Method B vs Method C

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Method B	12	61.25	5.104167	5.72964
Method C	12	53.625	4.46875	5.679332

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	2.422526	1	2.422526	0.42467	**0.521366**	4.300949
Within Groups	125.4987	22	5.704486			
Total	127.9212	23				

ANOVAs for Section 6.3

When Weasel exhibits brittle behavior, do some teams choose only Weasel's flawed solution set? Yes

Anova: Single Factor			**Team decisions in brittle scenario: Teams choosing flawed Weasel solution**			
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Not Weasel	12	2	0.166667	0.151515		
Chose Weasel	12	8	0.666667	0.242424		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	1.5	1	1.5	7.615385	**0.011436**	2.948585
Within Groups	4.333333	22	0.19697			
Total	5.833333	23				

Table above shows a significant difference in *team decisions* for the brittle scenario; teams chose Weasel's flawed solution set significantly more.

ANOVAs for Section 6.4

Does ECOA quality decline when Weasel exhibits brittle behavior? No

Anova: Single Factor	Indiv ECOA Quality Between Scen 1 and Scen 2					
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Scen 1 ECOA Qual	18	117.25	6.513889	2.190972		
Scen 2 ECOA Qual	18	105.875	5.881944	1.556015		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	3.594184	1	3.594184	1.918439	**0.17505**	2.859225
Within Groups	63.69878	34	1.873494			
Total	67.29297	35				

Table above shows no significant difference in individual's ECOA quality between scenario 1 and 2.

Anova: Single Factor	Teams ECOA Quality Between Scen 1 and 2					
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Scen 1 ECOA Qual	11	63.125	5.738636	3.442045		
Scen 2 ECOA Qual	12	62.375	5.197917	11.00414		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	1.677994	1	1.677994	0.22666	**0.638927**	2.960956
Within Groups	155.466	21	7.403144			
Total	157.144	22				

Table above shows no significant difference in team's ECOA quality between scenario 1 and 2.

ANOVAs for Section 6.5

Does Weasel help teams overall to produce better quality ECOAs? No

Team ECOA Quality Scores With (Method B/C) Weasel vs. Without (Method A) Weasel						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
ECOA Qual. Score w/out Weasel (Meth A or B w/ no revisions)	12	68.375	5.697917	3.975734		
Ave. ECOA Score w/ Weasel (Meth B w/ revisions and C)	9	57.125	6.347222	5.753038		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	2.16821	1	2.168217	0.458972	**0.506271**	2.9899
Within Groups	89.7573	19	4.724073			
Total	91.9255	20				

Table above shows no significant difference in ECOA quality scores with Weasel (method B/C) than without Weasel (method A).

ANOVAs for Section 6.6

Does Weasel help users working in teams better than individuals? No

Scenario 1:

Anova: Single Factor — Team vs Individual **Scores** for Scenario 1

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Indiv Scores	18	92.375	5.131944	6.464103
Team Scores	12	70.625	5.885417	2.887666

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	4.087587	1	4.087587	0.807971	**0.376386**	4.195972
Within Groups	141.6541	28	5.059074			
Total	145.7417	29				

Table above shows no significant difference in Scenario 1 decision quality scores between individuals and teams.

Anova: Single Factor — Team vs Individual **Ranks** for Scenario 1

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Indiv Ranks	18	292.5	16.25	86.44118
Team Ranks	12	172.5	14.375	61.26705

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	25.3125	1	25.3125	0.33066	**0.569863**	4.195972
Within Groups	2143.438	28	76.55134			
Total	2168.75	29				

Table above shows no significant difference in Scenario 1 rankings between individuals and teams.

Scenario 2:

Anova: Single Factor			Team vs Individual **Scores** for Scenario 2			
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Indiv Scores	18	85.125	4.729167	4.085018		
Team Scores	12	55.25	4.604167	6.391572		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	0.1125	1	0.1125	0.02254	**0.881736**	2.893846
Within Groups	139.7526	28	4.991164			
Total	139.8651	29				

Table above shows no significant difference in Scenario 2 decision quality scores between individuals and teams.

Anova: Single Factor			Team vs Individual **Ranks** for Scen 2			
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Indiv Ranks	18	276.5	15.36111	68.54575		
Team Ranks	12	188.5	15.70833	89.73674		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	0.868056	1	0.868056	0.011292	**0.916129**	2.893846
Within Groups	2152.382	28	76.87078			
Total	2153.25	29				

Table above shows no significant difference in Scenario 2 rankings between individuals and teams.

Overall:

Anova: Single Factor			Overall: Team vs Indiv **Scores**			
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Indiv Scores	18	177.5	9.861111	14.85274		
Team Scores	12	125.875	10.48958	13.1746		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	2.843837	1	2.843837	0.200362	**0.657873**	2.893846
Within Groups	397.4171	28	14.19347			
Total	400.2609	29				

Table above shows no significant difference in overall decision quality scores between individuals and teams.

Anova: Single Factor			Overall: Team vs Indiv **Ranks**			
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Indiv Ranks	18	281	15.61111	78.44281		
Team Ranks	12	184.5	15.375	76.03977		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	0.401389	1	0.401389	0.005179	**0.94314**	2.893846
Within Groups	2169.965	28	77.49876			
Total	2170.367	29				

Table above shows no significant difference in overall rankings between individuals and teams.

ANOVAs for Section 6.7

Does presentation order increase preference toward computer solutions? No

Anova: Single Factor — **Presentation Order vs Decision Made (how many picked computer set vs team set)**

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Method A	8	6	0.75	0.214286
Method C	8	4	0.5	0.285714

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	0.25	1	0.25	1	**0.334282**	4.60011
Within Groups	3.5	14	0.25			
Total	3.75	15				

Table above shows no significant difference between the decisions made (choosing computer or team ECOA set) and presentation order (method A: generate team ECOAs first vs method C: view computer ECOA set first).

Analysis on different decision behavior between teams and individuals:

Anova: Single Factor			**Overall: Team vs Indiv Choosing Computer Set Teams Chose Computer Set Signficantly More**			
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Indiv	36	8	0.222222	0.177778		
Teams	24	13	0.541667	0.259058		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	1.46944444	1	1.469444	6.997035	**0.010489**	4.006873
Within Groups	12.1805556	58	0.21001			
Total	13.65	59				

Table above shows significant difference in *overall decision choices* between teams and individuals; teams chose the computer ECOAs significantly more than individuals did.

Anova: Single Factor			**Overall - Teams vs Indiv Choosing Manual Set Indiv Chose Manual Set Signficantly More**			
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Indiv	36	18	0.5	0.257143		
Teams	24	4	0.166667	0.144928		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	1.6	1	1.6	7.524324	**0.008085**	4.006873
Within Groups	12.3333333	58	0.212644			
Total	13.9333333	59				

Table above shows significant difference in *overall decision choices* between teams and individuals; this analysis shows individuals chose their manually generated ECOAs significantly more than teams did.

Anova: Single Factor			**Scen 2 Indiv vs Team Decisions** **Teams Chose Computer Set More**			
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Individuals	18	5	0.277778	0.212418		
Teams	12	8	0.666667	0.242424		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	1.08888889	1	1.088889	4.856637	**0.035931**	4.195972
Within Groups	6.27777778	28	0.224206			
Total	7.36666667	29				

Table above shows significant difference in *scenario 2 decision choices* between teams and individuals; teams chose the computer ECOAs significantly more than individuals did.

Anova: Single Factor			Scenario 1 Indiv vs Team Decisions No significant differences			
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Individuals	18	3	0.166667	0.147059		
Teams	12	5	0.416667	0.265152		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	0.45	1	0.45	2.326154	**0.138433**	4.195972
Within Groups	5.41666667	28	0.193452			
Total	5.86666667	29				

Table above shows no significant difference in *scenario 1 decision choices* between teams and individuals.

ANOVAs for Section 6.8

Does order of presentation impact performance? No

Anova: Single Factor	**Presentation order vs ECOA quality scores**					
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Method B	8	48.4875	6.0609375	4.2506892		
Method C	6	29.375	4.895833333	8.0276042		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	4.65417	1	4.654175037	0.7990818	**0.3889448**	3.1765
Within Groups	69.8928	12	5.824403754			
Total	74.5470	13				

Table above shows no significant difference between team ECOA quality scores and presentation order (method B: generate manual ECOAs, view computer ECOAs, then revisions allowed vs Method C: view computer ECOAs, then generate manual ECOAs).

ANOVAs for Section 6.9

How do user attitudes (regarding communication, satisfaction, confidence, etc.) relate to performance or decision behavior?

Individuals vs. Teams:

Anova: Single Factor — **DM-Trust Question 4: Indiv vs Teams**
Attitude about computers (Indiv more favorable)

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Indiv	18	77	4.277778	0.683007
Teams	36	135	3.75	0.592857

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	3.342593	1	3.342593	5.371102	**0.024443**	2.804584
Within Groups	32.36111	52	0.622329			
Total	35.7037	53				

Table above shows significant difference in *attitudes about computers* between teams and individuals; individual participants had significantly more favorable attitudes.

Anova: Single Factor — **DM-Trust Question 5: Indiv vs Teams**
Hrs/week playing wargame sim (indiv more)

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Indiv	18	39	2.166667	1.088235
Teams	36	56	1.555556	1.111111

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	4.481481	1	4.481481	4.060665	**0.049076**	4.026631
Within Groups	57.38889	52	1.103632			
Total	61.87037	53				

Table above shows significant difference in *hours/week playing wargame* between teams and individuals; individual participants played significantly more.

Individuals vs. Teams:

Anova: Single Factor		**DM-Trust Question 7: Indiv vs Teams Confid in comp ID important ECOAs (teams more)**				
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Indiv	18	52	2.888889	0.457516		
Teams	36	129	3.583333	0.592857		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	5.787037	1	5.787037	10.54852	**0.002039**	2.804584
Within Groups	28.52778	52	0.548611			
Total	34.31481	53				

Table above shows significant difference in *confidence in computer identifying important ECOAs* between teams and individuals; teams were significantly more confident than individuals.

Team Leaders vs. Subordinates:

Anova: Single Factor		**DSS Rsch Spec Survey Item 2: Intel Quality Subs rated significantly higher**				
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
CC's	12	41	3.416667	0.628788		
Subs	24	95	3.958333	0.389493		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	2.347222	1	2.347222	5.027122	**0.031586**	2.859225
Within Groups	15.875	34	0.466912			
Total	18.22222	35				

Table above shows significant difference in *quality of computer intelligence ratings* between leaders and subordinates; subordinates rated intelligence quality significantly higher.

Team Leaders vs. Subordinates:

Anova: Single Factor		**DSS Rsch Spec Survey Item 9: Confidence in Tool** **Subs rated confidence significantly higher**				
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
CC's	12	39	3.25	0.75		
Subs	24	92	3.833333	0.405797		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	2.722222	1	2.722222	5.263823	**0.028074**	2.859225
Within Groups	17.58333	34	0.517157			
Total	20.30556	35				

Table above shows significant difference in *confidence in DSS tool* between leaders and subordinates; subordinates had significantly higher confidence.

Anova: Single Factor		**Team Interaction Item 4: Prefer leader to make dec** **Leaders rated significantly higher**				
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
CC's	12	49	4.083333	0.628788		
Subs	24	76	3.166667	1.362319		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	6.722222	1	6.722222	5.975309	**0.019846**	2.859225
Within Groups	38.25	34	1.125			
Total	44.97222	35				

Table above shows significant difference between leaders and subordinates in *preference for leader to make the final decision*; leaders rated preference to make decision significantly higher than subordinates.

Team Leaders vs. Subordinates:

Anova: Single Factor		**Team Interaction Item 9: Team Effective in Dec Quality** **Subs rated signif higher**				
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
CC's	12	51	4.25	0.386364		
Subs	24	112	4.666667	0.231884		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	1.388889	1	1.388889	4.927536	**0.033205**	2.859225
Within Groups	9.583333	34	0.281863			
Total	10.97222	35				

Table above shows significant difference in *ratings of team effectiveness in decision quality* between leaders and subordinates; subordinates rated effectiveness significantly higher than leaders.

Anova: Single Factor		**Team Interaction Item11: Team effective in working together** **Subs rated signif higher**				
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
CC's	12	52	4.333333	0.424242		
Subs	24	114	4.75	0.195652		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	1.388889	1	1.388889	5.151515	**0.029684**	2.859225
Within Groups	9.166667	34	0.269608			
Total	10.55556	35				

Table above shows significant difference in *ratings of team effectiveness in working together* between leaders and subordinates; subordinates rated team effectiveness significantly higher than leaders did.

ANOVAs for Section 6.10: Additional Analyses

Do teams who communicate more perform better than teams who demonstrate less communication? Yes

Communication Quantity: Top-Ranked Teams vs Bottom-Ranked Teams
Top teams communicated more

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Top 4	4	557	139.25	874.25
Bottom 4	4	328	82	365.3333

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	6555.125	1	6555.125	10.57633	**0.017419801**	3.775949
Within Groups	3718.75	6	619.7916			
Total	10273.875	7				

Table above shows significant difference in *amount of communication* between top-ranked teams and bottom-ranked teams; top-ranked teams communicated significantly more.

Appendix 3.3 Regression Analysis Calculations

The tables below show the linear regression analysis discussed in section 6.10.

Is there a relationship between the number of ECOAs generated and solution quality? Yes

SUMMARY OUTPUT		**Total ECOAs made vs Total Solution Quality Scores**				
Regression Statistics						
Multiple R	0.491636					
R Square	0.241706					
Adjusted R Square	0.214624					
Standard Error	5.573335					
Observations	30					
ANOVA	*df*	*SS*	*MS*	*F*	***Significance F***	
Regression	1	277.229	277.229	8.925004	**0.005794**	
Residual	28	869.7377	31.06206			
Total	29	1146.967				
	Coefficients	*Standard Error*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*
Intercept	0.550656	2.995237	0.183844	0.855461	-5.58481	6.686122
X Variable 1	**0.832238**	0.278576	2.987474	**0.005794**	0.261602	1.402875

Table above shows significant positive relationship between *total number of ECOAs generated by a user and total solution quality score*; generating more ECOAs positively related to higher quality scores (“X variable 1” in table corresponds with solution quality score).

SUMMARY OUTPUT		**WITHOUT OUTLIER: Regression Analysis on Ttotal ECOAs Generated vs Total Solution Quality Scores**				
Regression Statistics						
Multiple R	0.387634					
R Square	0.15026					
Adjusted R Square	0.118788					
Standard Error	3.332877					
Observations	29					
ANOVA						
	df	*SS*	*MS*	*F*	***Significance F***	
Regression	1	53.03475	53.03475	4.774435	**0.037739**	
Residual	27	299.9178	11.10807			
Total	28	352.9526				
	Coefficients	*Standard Error*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*
Intercept	6.920147	1.488995	4.647528	7.85E-05	3.864981	9.975313
Total ECOAs	0.368308	0.168558	2.185048	**0.037739**	0.022455	0.714161

This table shows the regression analysis performed without the “outlier” of team 8’s total number of ECOAs generated. Again, table above shows significant positive relationship between *total number of ECOAs generated by a user and total solution quality score*; generating more ECOAs positively related to higher quality scores (“X variable 1” in table corresponds with solution quality score).

SUMMARY OUTPUT		**Scenario 2**: **# ECOAs made vs ECOA Solution Quality**				
Regression Statistics						
Multiple R	0.554044					
R Square	0.306965					
Adjusted R Square	0.282214					
Standard Error	2.484877					
Observations	30					
ANOVA	*df*	*SS*	*MS*	*F*	***Significance F***	
Regression	1	76.57754	76.57754	12.402	**0.00149**	
Residual	28	172.8891	6.174612			
Total	29	249.4667				
	Coefficients	*Standard Error*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*
Intercept	1.004368	1.082774	0.927588	0.361549	-1.21359	3.22233
X Variable 1	**0.739939**	0.210112	3.521647	**0.00149**	0.309545	1.170333

Table above shows significant positive relationship between *number of ECOAs generated by a user in scenario 2 and solution quality score in scenario 2*; generating more ECOAs is positively related to higher quality scores ("X variable 1" in table corresponds with solution quality score).