The Impact of Computer Decision Support on Military Decision Making

A THESIS
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
MASTER OF SCIENCE

December 2004

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **05 JAN 2005**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**
4. TITLE AND SUBTITLE **The Impact of Computer Decision Support on Military Decision Making**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Minnesota Minneapolis**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**100**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Acknowledgements

The author would like to thank his wife for her unwavering love and support throughout this graduate school program.

I also express gratitude to Dr. Caroline Hayes for her help with this project. Your guidance has been invaluable and I thank you very much.

Appreciation is also offered to Dr. Tarald Kvalseth and Dr. Thomas Stoffregen for serving on the author's thesis examination committee.

Lastly, I would like to thank Air Force ROTC Detachment 415 and the Army ROTC Detachment at the University of Minnesota for their participation in this project. The personnel in these organizations were extremely helpful and I thoroughly enjoyed working with all of you.

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.

Abstract

This thesis work analyzed military personnel decision making and attitudes towards automation using a component of a decision support system tool called Weasel. The primary goal of this study is to determine how Weasel impacts user performance and behavior. The decisions people make in military situations plays a vital role in determining the success or failure of operations. Previous work has been conducted in transportation domains such as aviation and driving. However, little, if any, research has been conducted in a military domain. This study analyzes behavior and performance in a military context with military personnel solving three strategic problems. Specific challenges addressed by this work are Weasel's overall impact on user performance; Weasel's effect on expert and novice users; user performance when Weasel exhibits questionable behavior; and the effect order of information presentation has on behavior and performance. The results of this experiment will help researchers and military personnel interested in decision making and decision support systems to better understand the decisions people make when using computer support. Additionally, information may be gained regarding situations where computer support and automation use may degrade performance. Military strategists such as commanding officers, Air Force air battle managers, and Army plans officers may benefit from this work.

Table of Contents

List of Figures

Chapter 1 Introduction

This thesis describes an experiment analyzing the use of a decision support system tool, Weasel, and its effects on decision maker performance. Weasel is a decision support system (DSS) designed to assist military planning staff in generating courses of action (COAs) for ground forces. The primary goal of the study is to determine how Weasel impacts user's decision making performance. A secondary goal is to learn how decision support systems and user attitudes towards automation impact decision making. Specifically, are there situations in which DSSs can degrade decision making and what types of users are helped most by automation? Limited generalizations can be made from a single study, but by examining multiple studies, some generalizations can be drawn.

Decision support tools are beginning to become accepted in a wide variety of high-criticality decision making tasks, many with life and death consequences. Thus, it is important to understand the effect a DSS may have on human decision makers.

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 operating or monitoring the system. Previous work on decision making and automation systems focuses on systems such as transportation (aviation and driving). The military domain, and specifically Weasel, has not been analyzed in regards to the impact a DSS may have on military operator behavior and performance. There is limited knowledge regarding Weasel's current usefulness in military operations.

This work addresses these challenges by testing military personnel decision making in three problem solving scenarios. The study also gathers 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. Specific questions addressed are:

- Does Weasel help users overall to produce better quality COAs?
- Does Weasel help novices more than experts?
- When Weasel exhibits brittle behavior, do some subjects 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?
- Do questionnaire responses provide insight into subject performance or decisions?

Answering these questions will provide insight into the usefulness of Weasel and increase understanding of DSS user behavior and performance in a military context. The significance of this research may be seen in improved military 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 knowledge in the areas of decision making and use of automation.

Chapter 2 Literature Review

Analyzing the types of decisions people make when working with automated systems requires an important understanding of the level of trust that individual have in the system being used and their overall trust in automation. Multiple authors have found various factors that lead to trust in automation, including automation reliability [7], operator attitudes [9], workload and situational risk [15]. However, the same factors have not been consistently found in all studies. Parasuraman [12] 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."

Even though automated tools can be of significant help in many situations, research shows systems that lack reliability can lead to high levels of disuse [21] and misuse [12] by operators. General findings are that users typically trust automated aids initially and trust wanes after some failure has occurred [6, 18]. Whether or not reliability is an issue, the decision to use or not use automated aids leads to some interesting tactics by humans. Vicente [19] says automation users, particularly novices, may adopt a strategy where they simply do what is easiest and not what may lead to better outcomes. If the automation system being used lacks quality, the result may lack quality as well. Users may also implement a strategy to maximize their correctness in working with automated decision aids by always agreeing with the aid, even though they knew there would be instances in which they were incorrect [21]. In the context of military operations, this maximization strategy could lead to catastrophic results. Events that are not diagnosed correctly may lead to devastating casualties or other types of military losses. Parasuraman and Riley [12] 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.

To help reduce the frequency and consequences of operator error in complex systems, decision support systems (DSS) have been developed [21]. Computer decision making and diagnostic aids can help reduce human error but as with most technology, there is a reliance on the operators themselves to correctly use, or not use, the system. These DSS tools often aid the user in balancing the large amounts of complex information mentioned in the introduction. Information is usually a good thing when making decisions, especially when those decisions affect the outcome of military battles. However, as the old adage says, more information isn't necessarily good information. As long as the information is accurate and useful, more typically is better. The decision theorist Thomas Cowan states when conflict is involved, "information is armament" if the information is deemed to be "good" [5]. Most people agree computers can typically store and process more information than humans, but that may not be necessarily better when it comes to decision making. There are other factors, such as instinct, stress, and real-world experience that humans can account for and computers often cannot. Without the experience and gut instincts of many fantastic military leaders over the years, we may not be afforded the privilege of living the lives we have.

Lee and Moray [8] 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 [2]. 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. This highlights the importance of properly training users in operating automated systems. A teenager driving an automobile for the first time typically is less comfortable than their parents who have driven for many years. Driver education classes and hours of driving experience make the teenager more confident at the wheel, sometimes to a fault. Likewise, subjects in research experiments must be trained to competently understand and use the system involved. When possible, instruction should include describing how the system operates while also providing experience using the automation as well [6]. Research has also shown the more complex a task is the more reliant people may be on the system, despite their experience and level of confidence [6].

Self-confidence is another key factor in determining user's perceived ability to control an automated system. Lee and Moray [8] 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 [8] but users can also be overconfident in automated system capabilities and accuracy. An example was when the Royal Majesty cruise ship ran aground near Nantucket after veering off course. The National Transportation Safety Board (NTSB) reported the accident was primarily caused by crew over-reliance on the ship's Automatic Radar Plotting Aid and Global Positioning System [11].

Trust in automation literature has highlighted many valuable findings to date. Further research analyzing trust and confidence in automation versus a person's own capabilities may provide additional insight into the type of decisions people make.

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 project offered subjects the opportunity to decide between automated actions, personal actions, and sometimes 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 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 with their own eyes. These sources of intelligence don't always concur with each other one hundred percent. The challenging task of the commander is to decipher which intelligence they feel is best and advise troops accordingly.

Humans are very adept at taking complex situations with many factors dependent highly on specific situational factors and accounting for the variables involved [16]. 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 project 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 [4] 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 [4]. 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 [4]. 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.

Confidence varies among humans and expert and novice subjects are no different. Interestingly enough research has shown trends where overconfidence increases as experience decreases [13]. This may counter what a person would logically think. The possibility that novice subjects simply don’t know what they don’t know may lead them to be overconfident in their abilities. In a complex task this overconfidence may decrease significantly after a short time and the novice realizes the complexity of the task.

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 [17] 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 project analyzed decisions made 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 [8]. A situation with ambiguity leaves the decision susceptible to human factors such as an individual's expectations and motivations [18]. It is important for researchers to recognize and identify any types of bias that may be evident in a subject's decision making process.

Chapter 3 Weasel: A Decision Support System

Weasel is one component of a decision support system (DSS) developed in 2003 at the University of Minnesota. It is programmed in C++ and uses a genetic algorithm (GA) to generate and evaluate plans of enemy and friendly military forces [16]. 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 are shown. This experiment utilized two components, the Weasel ECOA generator and the map display. Screen captures of the Weasel and map pages are shown in Figures 1 and 2 respectively.

The goals of Weasel were to 1) assist users by identifying a good set of likely enemy actions, 2) present high quality and diverse courses of action for users to select from, and 3) quickly generate and evaluate possible actions under time constraints [16]. Weasel is human-guided and constraint-based [14] as potential enemy actions are generated by the FOX-GA according to the information (i.e. possible intelligence data) entered into the computer and the applicable constraints of the situation.

The map display (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).

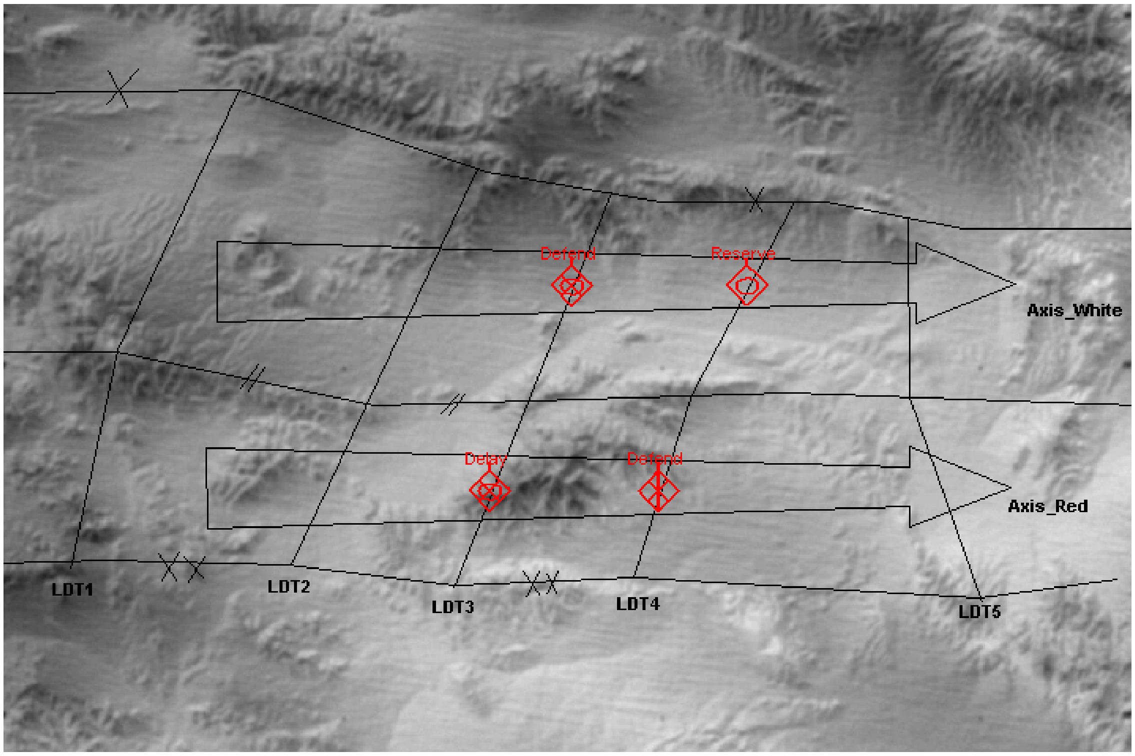


Figure 1. Map display example with enemy forces depicted

An avenue of approach (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 lines of defensible terrain (LDT). 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 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 be designated to defend (Def), delay in defense (Del), or defend in reserve (R).

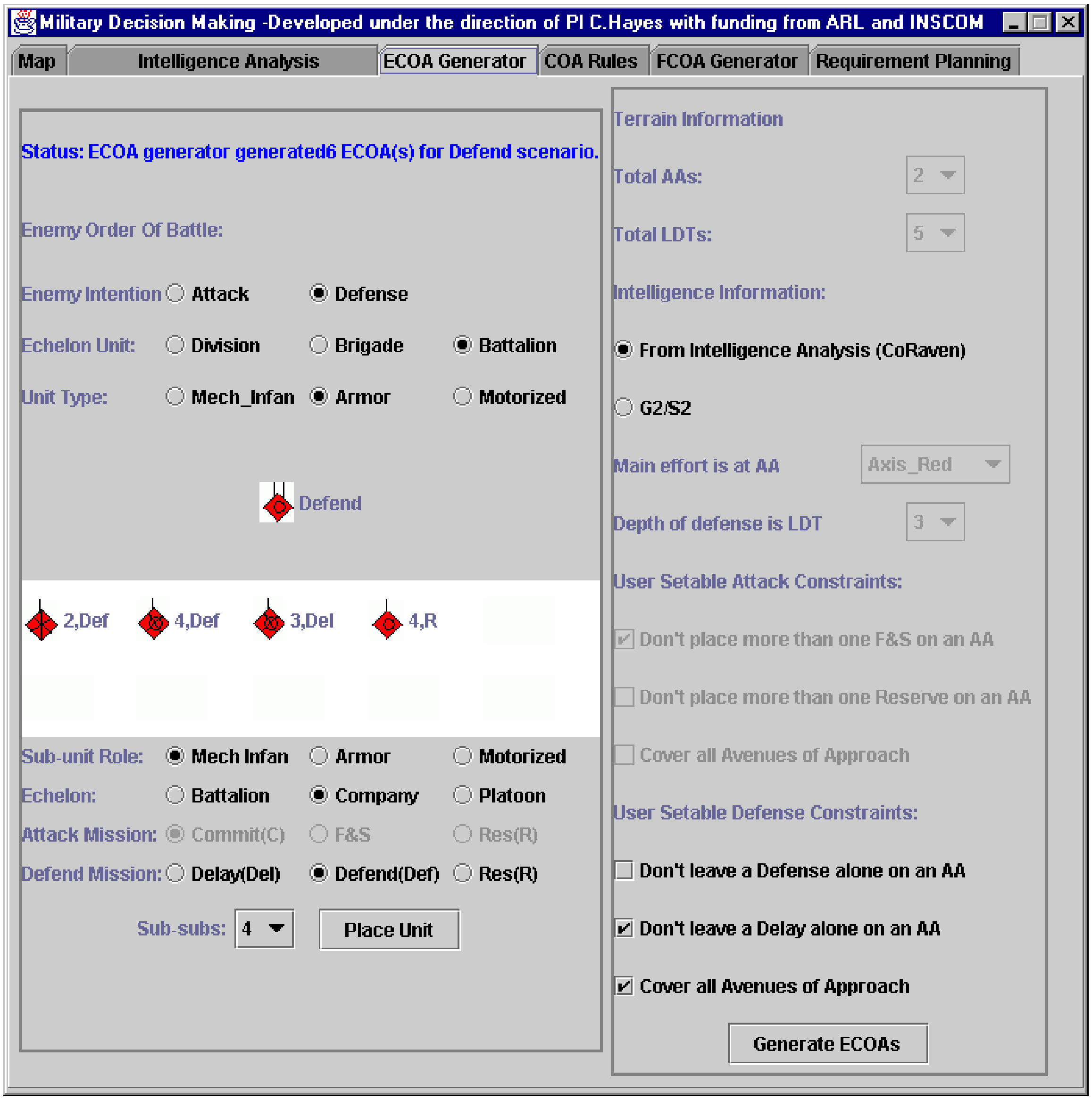


Figure 2. Input screen for Weasel ECOA generator

Once all enemy data or "intelligence" is entered 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.

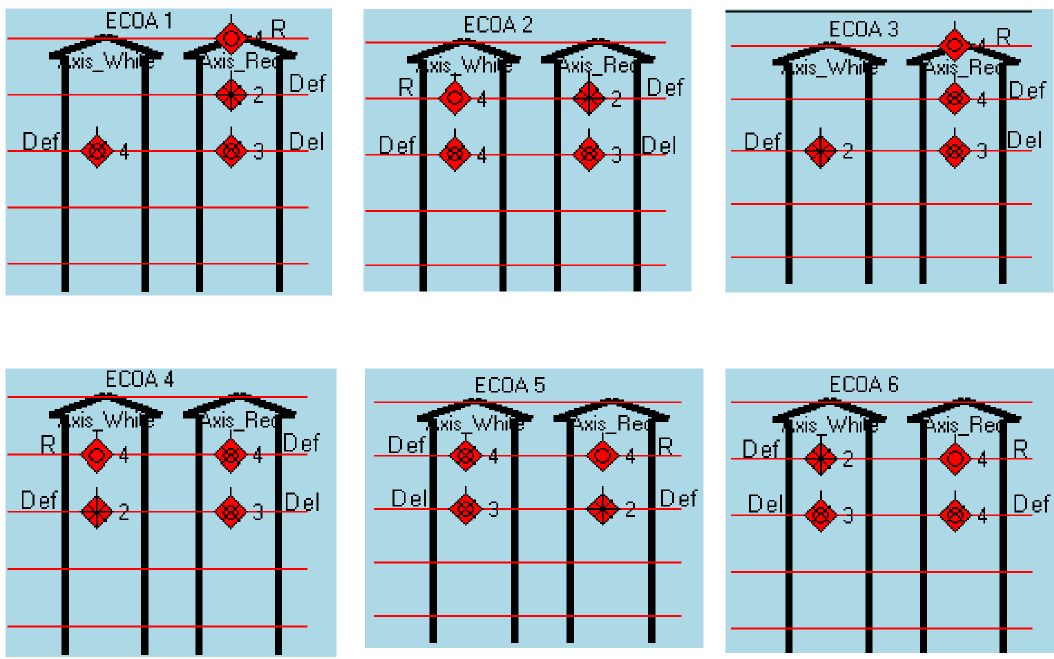


Figure 3. Example of ECOA set generated by Weasel

The U.S. Army symbology [19] used in the DSS to depict enemy force information is shown in the following figure.

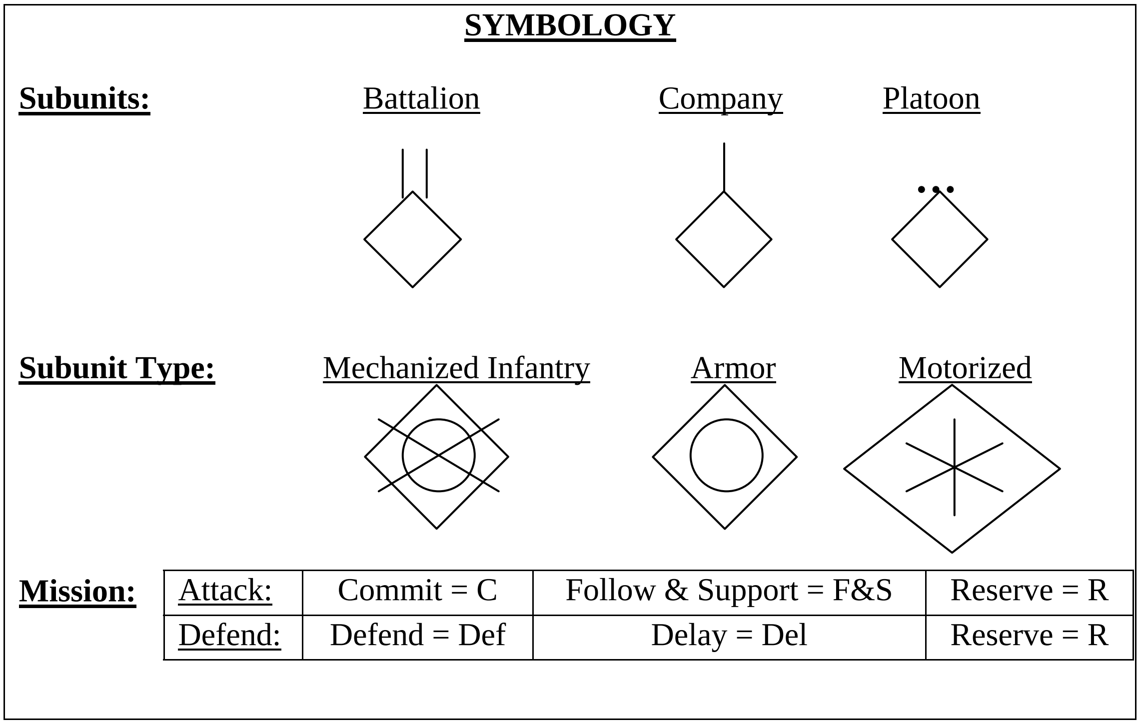


Figure 4. Unit symbology

An ECOA set generated by Weasel can be viewed in the FCOA generator tab in the portion highlighted in Figure 5. The courses of action shown in blue in Figure 5 are potential FCOAs.

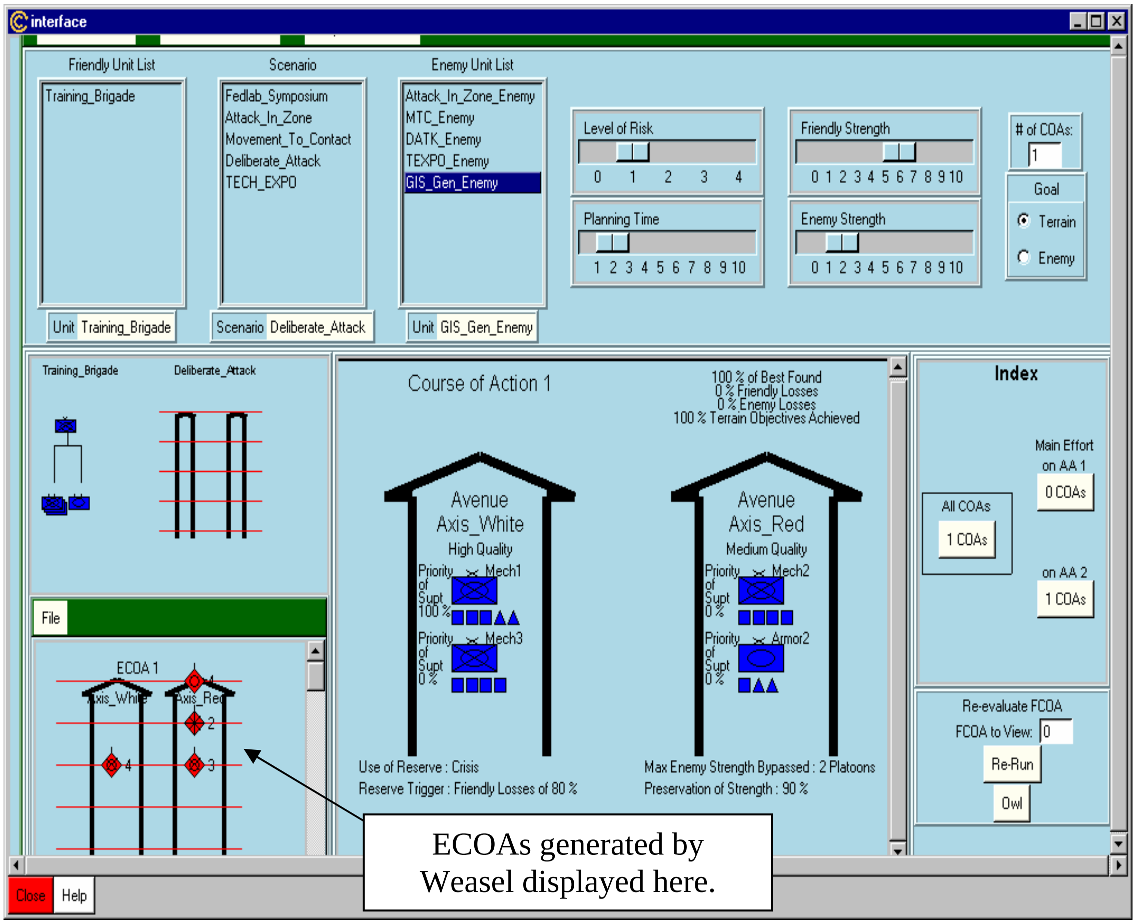


Figure 5. FCOA page highlighting ECOA generated

Chapter 4 Experiment Description

4.1 Subjects

There were 18 subjects that participated in the experiment (9 Air Force and 9 Army subjects). All had experience in the U.S. armed forces in addition to Officer Training School, Reserve Officer Training Corp (ROTC), Service Academy, or enlisted program training. Subjects were categorized as expert or novice; experts were those having at least 6 years of military experience on active duty, in the National Guard or Reserves. Novice subjects were current members of their respective service's Reserve Officer Training Corp (ROTC) program. The average length of experience of all 18 subjects was 5.03 years. Subjects were paid $10 per hour for participating. A summary of subject data is shown below in Figure 6. A complete listing of subject data can be seen in Figure 7.

Subject Breakdown

	Novice	Expert	Total
Air Force	5	4	9
Army	8	1	9
Total	13	5	18

Average Experience (yrs)

	Novice	Expert	Overall
Air Force	2.00	13.125	6.94
Army	2.50	8	3.11
Overall	2.31	12.10	5.03

Figure 6. Summary of subject information

Subject	Service	Experience Level	Years of Experience	Experience Description
1	AF	Novice	1	ROTC training
2	AF	Novice	2	ROTC training
4	AF	Novice	3	ROTC training
6	Army	Novice	3	ROTC training
7	Army	Novice	1	ROTC training
8	AF	Novice	1	ROTC training
9	Army	Novice	1	ROTC training
14	Army	Novice	3	ROTC training
18	AF	Novice	3	ROTC training
3	Army	Novice	3	1 year in National Guard and 2 years of ROTC training
5	Army	Novice	2	1 year in National Guard and 1 year of ROTC training
10	Army	Novice	3	Active duty 1 year and 2 years of ROTC training
12	Army	Novice	4	2 years in Army Reserves and National Guard and 2 years of ROTC training
11	AF	Expert	7.5	Active duty for 7.5 years, last 2 years concurrent with ROTC training
13	Army	Expert	8	8 years in the National Guard, last 2 years concurrent with ROTC training
15	AF	Expert	21	Retired, active duty for 21 years
16	AF	Expert	9	Commissioned officer for 7 years and two years of ROTC training
17	AF	Expert	15	Active duty for 15 years

Figure 7. Detailed subject information

4.2 Evaluators

Two evaluators participated in the experiment. They were selected because of their overall expertise regarding Army battlefield strategy as well as their specific knowledge of current battlefield simulations used in the U.S. Army. They have over 14 years of total U.S. Army experience between them.

Evaluator one is a U.S. Army officer who for 3 years served as a plans officer on General staff and was a platoon leader. He received a business administration degree from the University of Minnesota and is currently serving as an Assistant Professor of Military Science at the University of Minnesota in Minneapolis, Minnesota. Evaluator number 2 is also a commissioned U.S. Army officer who was serving in the Army ROTC detachment at the University of Minnesota at the time of the experiment. He is an infantry officer with 3 years of experience in planning maneuver strategies. He received a bachelor's degree in international relations from the University of Minnesota and received 2 years of extensive leadership and decision making training from the U.S. Army via infantry school, basic training, and leadership classes.

The evaluators have experience using major Army simulation systems which were helpful in their understanding of the automation used in this experiment. Two systems, the Battle Simulation Network (SIMNET) and Close Combat Tactical Trainer (CCTT), are platoon, company, and battalion maneuver training tools used for strategic training of military personnel.

4.3 Problem Scenarios

Three scenarios were derived for use in subject trials. Complete descriptions for scenarios 1, 2, and 3 can be found in Appendix 2.

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 on the decision maker. Scenarios 1 and 3 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 8.

<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.

Figure 8. Problem example – Scenario 1

Scenario 2 is the most basic of the scenarios devised. 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.

Scenario 3 was designed to elicit "brittle" [17] behavior from Weasel, meaning the set of solutions generated by the automated tool were lacking in quality. The intent for examining a brittle scenario was to analyze issues such as subject decisions and the quality of ECOAs generated; which may provide insight into the affect on subject performance when the DSS exhibits brittle behavior. The quality of the automated set was questionable due to the fact that one AA was always left open in each automated ECOA. This leaves an unrealistic opening for counter maneuvers by the opposing force. Scenario 3 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.

4.4 Experiment Design

Each subject completed the experiment by themselves with the researcher present at all times. A lattice [10] experiment design was used and is shown in Figure 9. This design eliminated any learning effects that may have occurred across the three scenarios and affords the opportunity to analyze for bias effects on the order of information presented (manual versus automated ECOAs presented to the subject first or second).

Scen. Order	ECOA Generation Method and Scenario Pairings Used By Subject					
	A1, B2, C3	**A1, C2, B3**	**B1, A2, C3**	**B1, C2, A3**	**C1, A2, B3**	**C1, B2, A3**
123		A1, C2, B3		B1, C2, A3		C1, B2, A3
132	A1, C3, B2		B1, C3, A2		C1, B3, A2	
231		C2, B3, A1		C2, A3, B1		B2, A3, C1
213	B2, A1, C3		A2, B1, C3		A2, C1, B3	
312		B3, A1, C2		A3, C2, B1		A3, C1, B2
321	C3, B2, A1		C3, A2, B1		B3, A2, C1	

Top row shows the ECOA generation and analysis method used with each scenario	**A** = manual ECOA 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	**3** = scenario 3

Figure 9. Experiment design matrix

The design matrix displays the presentation of scenarios for 18 subjects. The top row is the ECOA generation and analysis method to be used for three scenarios. The far-left column shows the order in which the subject would complete the scenarios. Each row and column had novice and expert subject representation.

In addition to the 18 subjects used for the research trials and data collection, five separate subjects were run through the experiment beforehand. The purpose of these practice trials was to work out any problems in the experiment and allow the researcher to gain a feel for the time involved in running each subject. The five practice trial subjects were comprised of 4 novices and 1 expert, all of whom were U.S. Air Force personnel.

4.5 Problem Solving Methods

Three problem solving methods were used by each subject. A different method was used to solve each of the three scenarios. Method A consisted of manual generation of ECOAs first, given ECOA set generated by Weasel second, analysis of the two sets, followed by deciding between the two sets as to which one was better. Revisions and mixing ECOAs in the final decision were not allowed. In method B, subjects again manually generated their own ECOAs first, were then given the computer generated ECOAs, at this time revisions could be made to their manual set, and then they made their decision regarding which ECOAs were best. Choosing from both ECOA sets was allowed in method B. In method C, subjects were shown the ECOAs from Weasel first. They generated manual ECOAs next as well as being allowed to make any revisions deemed necessary. Lastly, they choose their ideal ECOA set and again, mixing from both sets was allowed in method C.

4.6 Subject Procedure

Subjects first read a brief explanation of the study (provided in Appendix 2.6). In this explanation the subject task for experimental trials was 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 then conducted by the researcher on a computer workstation in the DSS laboratory where the entire experiment was accomplished.

Training consisted of an in-depth explanation and discussion of the ECOA and FCOA generator tools, map display, constraint list, and applicable terms and symbols. Subjects were permitted to ask any questions at all times during the training and subjects were encouraged to use the DSS tool for hands-on experience at any time. The hard copy printouts of the constraints and acronym and symbology page were provided during the training. Subjects were provided as much time as they needed to study the symbology and constraints before officially starting the scenarios. Training was completed when all pertinent information had been explained and the subject verbally acknowledged they felt comfortable in understanding the experiment, task, and applicable tools.

The necessary materials for subjects to complete the scenarios were a scenario instruction page, the three pages describing one scenario per page, 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 2.7. All materials were provided by the researcher.

Before starting the scenarios all subjects read the page of instructions shown below in Figure 10. Each subject 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. You will have a limit of 30 minutes, TOTAL, to work on each scenario.

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

Figure 10. Scenario instructions

Subjects then completed each of the three scenarios. Subjects were allowed to ask questions pertaining to general understanding during this time. When the subject acknowledged 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. Subject comments and decisions were recorded by the researcher on a data recording sheet.

Upon completion of the three scenarios subjects completed a short questionnaire and were then released.

4.7 Evaluator Procedure

A critical aspect of the experiment was evaluation of the scenarios, ECOAs generated by the computer and subjects, and the decisions made by the 18 subjects. One expert evaluator assessed three scenarios prior to experimental trials being run to make sure they were logical, real-world situations that could arise in military operations. The same evaluator and one other expert each analyzed the quality of the ECOAs generated and the decisions subjects made in each scenario.

The evaluators underwent familiarization training on the DSS tool prior to evaluating any subject performance. The Weasel and map display tools were the focus of training, along with constraints and terminology.

All data given to evaluators was in paper format. The data provided to the evaluators was the three scenario descriptions (including any notes the subject may have written on the scenario pages), all maps with ECOAs generated by the subject, and the decision made by each subject 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 subjects, and the list of constraints applied in the DSS simulation.

Evaluators were instructed to evaluate subject performance in each scenario by analyzing the quality of the ECOAs generated (personal and automated) and the decision made by the subject versus the evaluator's ideal solution for the specific scenario. Subjects were ranked 1-18 on each scenario based on a 10 point scale and an overall performance ranking was obtained by averaging the ranks from the two evaluators from the three scenarios.

Evaluators analyzed the data for one week. Subject rankings and scores were recorded by the evaluators in spreadsheet format and they provided a hard copy of the rankings to the researcher. The results will be discussed in the following chapters.

4.8 Data Recorded

The following data was gathered on each subject during this experiment:

- Subject name
- Data user completed experiment
- Subject number assigned according to sequence of completion
- Subject's military service branch
- Subject's military experience (years and type of experience)
- Time spent completing familiarization training
- Time spent on each scenario
- Handwritten personal ECOAs generated by each subject
- Number of personal ECOAs generated for each scenario
- Number of personal ECOAs revised for each scenario (when revisions permitted)
- Decision on best ECOA set for each scenario
- Verbal comments for each scenario
- Scenario sequence and order of ECOA presentation or analysis
- Subject questionnaire upon completion of all scenarios

The exact questionnaire items asked of each subject is shown in Appendix 2.4. Example items include rating general trust in computers, confidence in identifying ECOAs, and confidence in the decisions made. The questionnaire rating scales follow the linear numeric design [1] and are commonly used in the Air Force.

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

Chapter 5 Results

5.1 Subject Solutions

Subject's choices (computer ECOAs vs. manual ECOAs or both) are shown in Figure 11. The table depicts the decision and number of ECOAs the subject generated for each scenario as well as the subject's experience level and military service branch. Data is sorted according to level of experience (novice or expert).

Subject #	Years of Experience	Brach of Service	Experience Level	Scenario 1 Decision	Scen 1 # of manual ECOA generated	Scenario 2 Decision	Scen 2 # of manual ECOA generated	Scenario 3 Decision	Scen 3 # of manual ECOA generated
1	1	AF	Novice	Manual	5	Manual	7	Manual	6
7	1	Army	Novice	Both	6	Manual	6	Both	6
8	1	AF	Novice	Automated	8	Manual	6	Manual	8
9	1	Army	Novice	Automated	4	Manual	4	Automated	6
2	2	AF	Novice	Manual	5	Manual	5	Both	5
5	2	Army	Novice	Manual	5	Manual	4	Both	6
4	3	AF	Novice	Manual	5	Manual	4	Automated	4
6	3	Army	Novice	Both	3	Manual	4	Manual	3
14	3	Army	Novice	Both	4	Manual	5	Manual	4
18	3	AF	Novice	Manual	5	Both	1	Both	6
3	3	Army	Novice	Manual	3	Both	2	Manual	6
10	3	Army	Novice	Both	2	Manual	3	Manual	3
12	4	Army	Novice	Both	6	Both	4	Manual	4
11	7.5	AF	Expert	Manual	3	Manual	5	Manual	4
13	8	Army	Expert	Automated	1	Automated	0	Manual	3
16	9	AF	Expert	Both	3	Both	1	Automated	2
17	15	AF	Expert	Manual	2	Automated	2	Automated	0
15	21	AF	Expert	Manual	2	Manual	4	Automated	1

Figure 11. Subject solutions

Shown in Figure 12 is summary data of the decisions made by subjects.

Number of Subjects Choosing Each Option Across Scenarios							
Scenario 1		**Scenario 2**		**Scenario 3**		**Totals**	
Manual	9	Manual	12	Manual	9	Manual	30
Automated	3	Automated	2	Automated	5	Automated	10
Both	6	Both	4	Both	4	Both	14
Totals	18		18		18		54

Figure 12. Summary of subject decisions

5.2 Evaluator Rankings

5.2.1 Scenario 1 Rankings

Figure 13 displays the evaluator rankings and solution quality scores for subjects in Scenario 1 grouped by experience level. Included is the score each subject received on the 10-point subjective solution quality rating, where a 10 means the subject perfectly matched the ideal solution derived by the evaluator, and their corresponding rank for scenario 1. The best rank = 1 and the worst = 18. Rankings were based on the score each subject received on a 10-point subjective solution quality rating by each evaluator. 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 18 was assigned to the subject with the lowest quality rating on the 10-point scale.

Years Experience	Experience Group	Evaluator 1		Evaluator 2		Ave. of Eval 1 & 2	
		Rank	**Score**	**Rank**	**Score**	**Rank**	**Score**
1	Novice	3.5	7.5	2	8.25	2.75	7.875
1	Novice	3.5	7.5	4	7.75	3.75	7.625
1	Novice	11	6.75	12	6.25	11.5	6.5
1	Novice	12	6.25	13.5	6	12.75	6.125
2	Novice	1	8.25	1	9	1	8.625
2	Novice	8	7	3	8	5.5	7.5
3	Novice	8	7	6.5	7.25	7.25	7.125
3	Novice	8	7	9.5	6.75	8.75	6.875
3	Novice	15	5.5	15	5.75	15	5.625
3	Novice	16	5.25	17	4.5	16.5	4.875
3	Novice	13.5	6	11	6.5	12.25	6.25
3	Novice	13.5	6	13.5	6	13.5	6
4	Novice	5	7.25	6.5	7.25	5.75	7.25
7.5	Expert	17	5	16	5.5	16.5	5.25
8	Expert	8	7	5	7.5	6.5	7.25
9	Expert	8	7	9.5	6.75	8.75	6.875
15	Expert	18	2	18	2	18	2
21	Expert	2	8	8	7	5	7.5

Figure 13. Scenario 1 rankings

5.2.2 Scenario 2 Rankings

Figure 14 displays the evaluator rankings and solution quality scores for subjects in Scenario 2 grouped by experience level.

Years Experience	Experience Group	Evaluator 1 Rank	Evaluator 1 Score	Evaluator 2 Rank	Evaluator 2 Score	Ave. of Eval 1 & 2 Rank	Ave. of Eval 1 & 2 Score
1	Novice	1	8.5	1	9.5	1	9
1	Novice	7	7	5	7.5	6	7.25
1	Novice	9	6.75	7.5	7	8.25	6.875
1	Novice	17.5	3	18	2.5	17.75	2.75
2	Novice	16	5.25	15	5	15.5	5.125
2	Novice	11.5	6.5	10.5	6.5	11	6.5
3	Novice	3	8	2	9	2.5	8.5
3	Novice	3	8	4	7.75	3.5	7.875
3	Novice	5	7.25	7.5	7	6.25	7.125
3	Novice	11.5	6.5	12	6.25	11.75	6.375
3	Novice	11.5	6.5	10.5	6.5	11	6.5
3	Novice	15	6	16	3.5	15.5	4.75
4	Novice	11.5	6.5	14	5.5	12.75	6
7.5	Expert	14	6.25	13	6	13.5	6.125
8	Expert	7	7	9	6.75	8	6.875
9	Expert	3	8	3	8	3	8
15	Expert	7	7	6	7.25	6.5	7.125
21	Expert	17.5	3	17	3	17.25	3

Figure 14. Scenario 2 rankings

5.2.3 Scenario 3 Rankings

Figure 15 displays the evaluator rankings and solution quality scores for subjects in Scenario 3 grouped by experience level and years of experience.

Years Experience	**Experience Group**	**Evaluator 1**		**Evaluator 2**		**Ave. of Eval 1 & 2**	
		Rank	**Score**	**Rank**	**Score**	**Rank**	**Score**
1	Novice	7.5	6	8.5	6.25	8	6.125
1	Novice	1	8	1	8	1	8
1	Novice	6	6.25	10	6	8	6.125
1	Novice	2	7.5	5	6.75	3.5	7.125
2	Novice	3	7	3	7.25	3	7.125
2	Novice	9	5.75	8.5	6.25	8.75	6
3	Novice	4.5	6.5	4	7	4.25	6.75
3	Novice	17	4.5	17	4	17	4.25
3	Novice	7.5	6	6.5	6.5	7	6.25
3	Novice	13	5.25	13	5.25	13	5.25
3	Novice	15	5	12	5.5	13.5	5.25
3	Novice	16	4.75	15	4.5	15.5	4.625
4	Novice	4.5	6.5	2	7.5	3.25	7
7.5	Expert	13	5.25	6.5	6.5	9.75	5.875
8	Expert	13	5.25	11	5.75	12	5.5
9	Expert	18	3	14	5	16	4
15	Expert	10.5	5.5	16	4.25	13.25	4.875
21	Expert	10.5	5.5	18	3.5	14.25	4.5

Figure 15. Scenario 3 rankings

5.2.4 Overall Rankings

Provided in Figure 16 are the overall rankings based on the average of the rankings each subject received from the two evaluators in each of the three scenarios. Included in the table are the subject's years of experience, experience group, average ranking from the evaluators, and overall ranking for the 3 scenarios. The best rank is 1 through the worst ranking of 18.

Years Experience	Experience Group	Branch of Service	Average Rank: All Scenarios	Overall Rank
1	Novice	AF	4.67	1
1	Novice	Army	7.25	4.5
1	Novice	AF	7.25	4.5
1	Novice	Army	12.83	16
2	Novice	AF	5.83	3
2	Novice	Army	10.00	11
3	Novice	AF	5.67	2
3	Novice	Army	8.67	7
3	Novice	Army	9.50	10
3	Novice	AF	10.50	12
3	Novice	Army	10.67	13
3	Novice	Army	14.83	18
4	Novice	Army	9.42	9
7.5	Expert	AF	11.08	14
8	Expert	Army	8.83	8
9	Expert	AF	8.33	6
15	Expert	AF	12.92	17
21	Expert	AF	12.75	15

Figure 16. Overall rankings across all scenarios (1 = best; 18 = worst)

5.3 Survey Responses

The data collected via the survey is provided below in Figure 17. The same questions were asked of both novice and expert subjects. The questionnaire can be seen in Appendix 2.4. Also shown are graphical depictions of subject responses for each question versus years of experience.

Some descriptions regarding 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.

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
Novice Subject 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
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
12	4	3	4	5	2	4	4	4	human
14	3	4	4	4	1	3	3	3	human
18	3	2	5	3	2	3	2	4	human
Expert Subject Responses									
11	4	3	4	4	1	4	4	4	human
13	5	4	3	4	3	3	4	3	computer
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

Figure 17. Subject survey responses

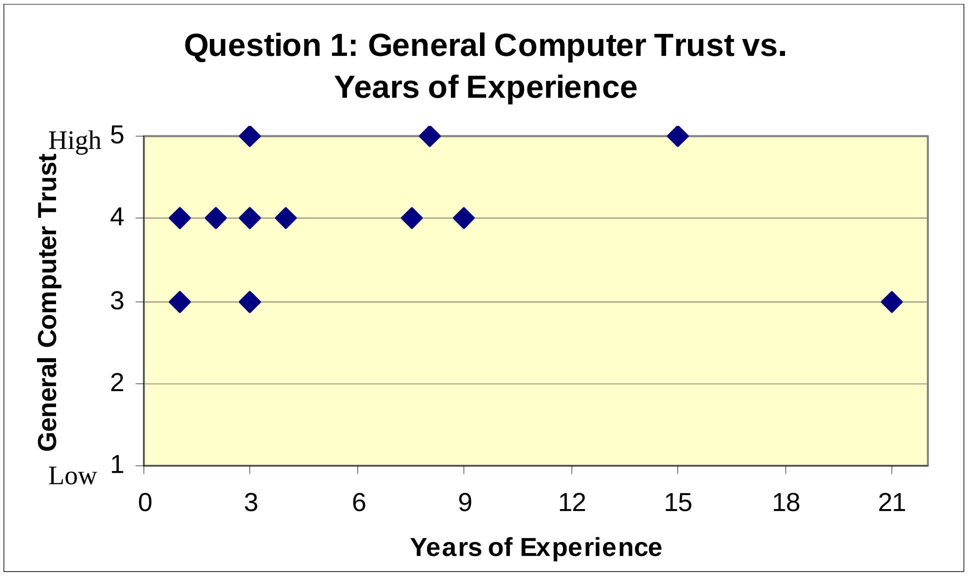


Figure 18. Questionnaire item 1 graph – general computer trust

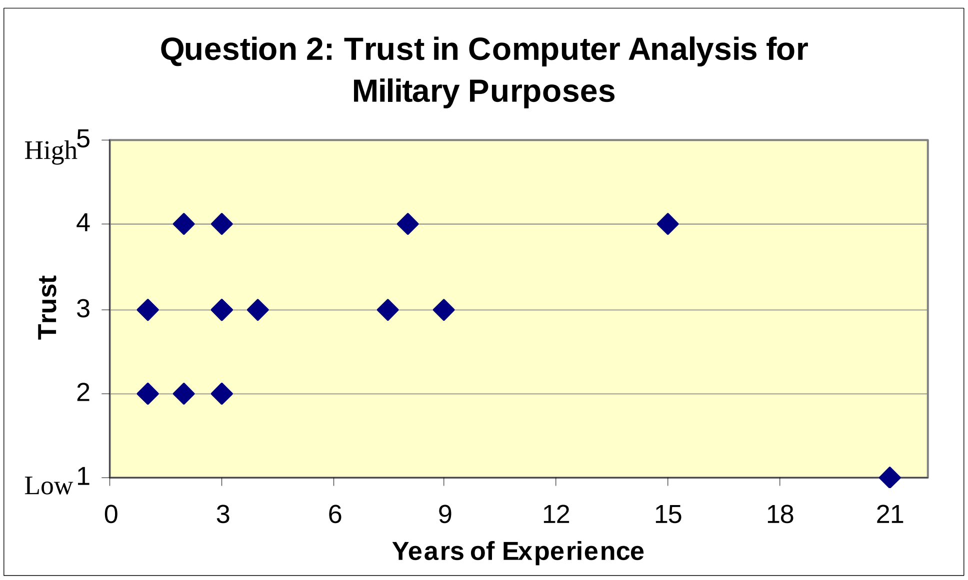


Figure 19. Questionnaire item 2 graph – trust in computer analysis

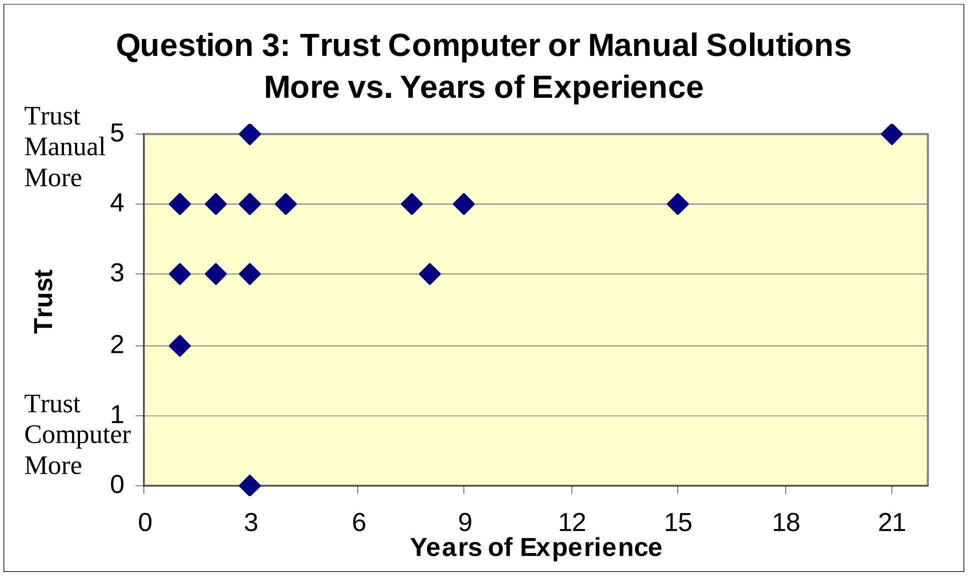


Figure 20. Questionnaire item 3 graph – trust in computer or manual solutions (0 = insufficient info to answer)

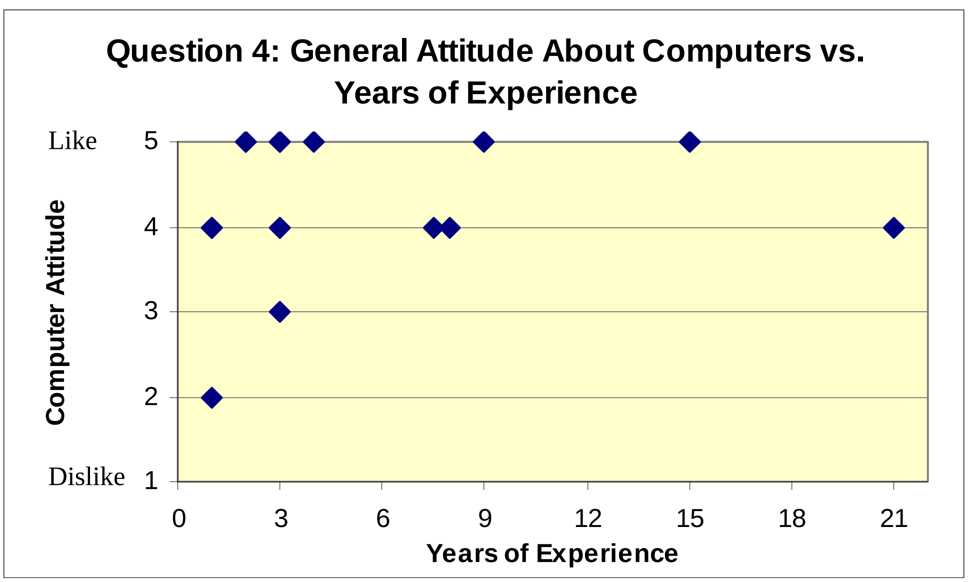


Figure 21. Questionnaire item 4 graph –attitude toward computers

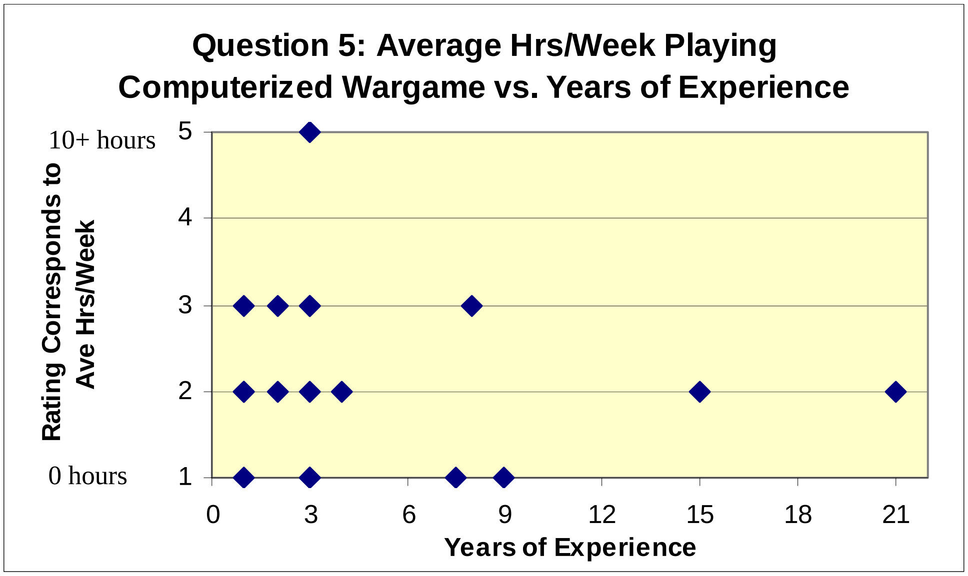


Figure 22. Questionnaire item 5 graph – wargame experience

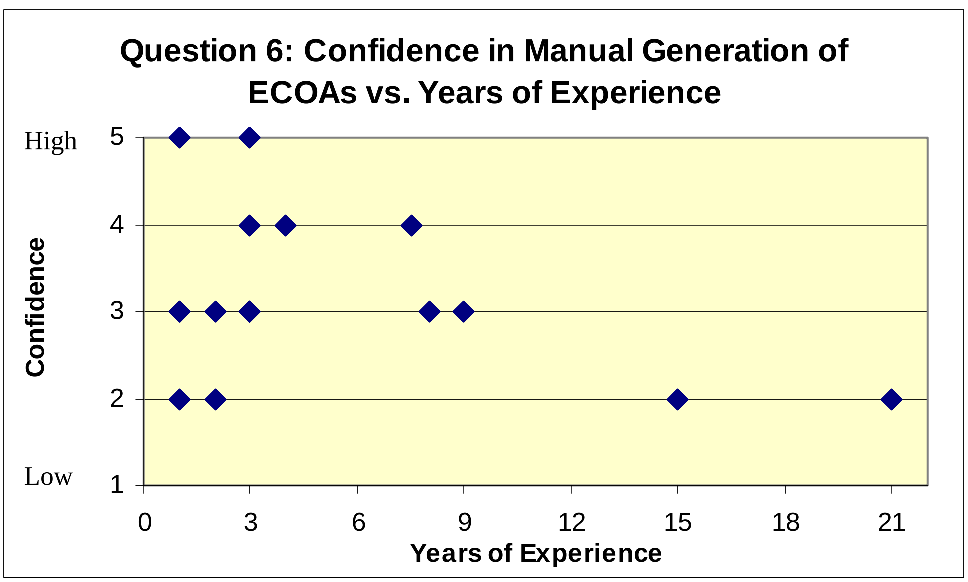


Figure 23. Questionnaire item 6 graph – manual ECOA confidence

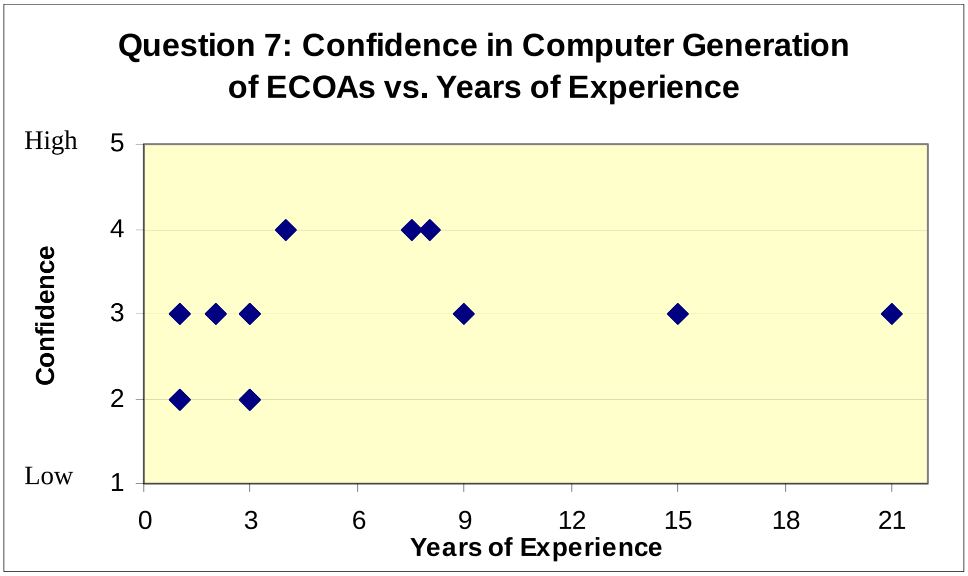


Figure 24. Questionnaire item 7 graph – confidence in Weasel ECOAs

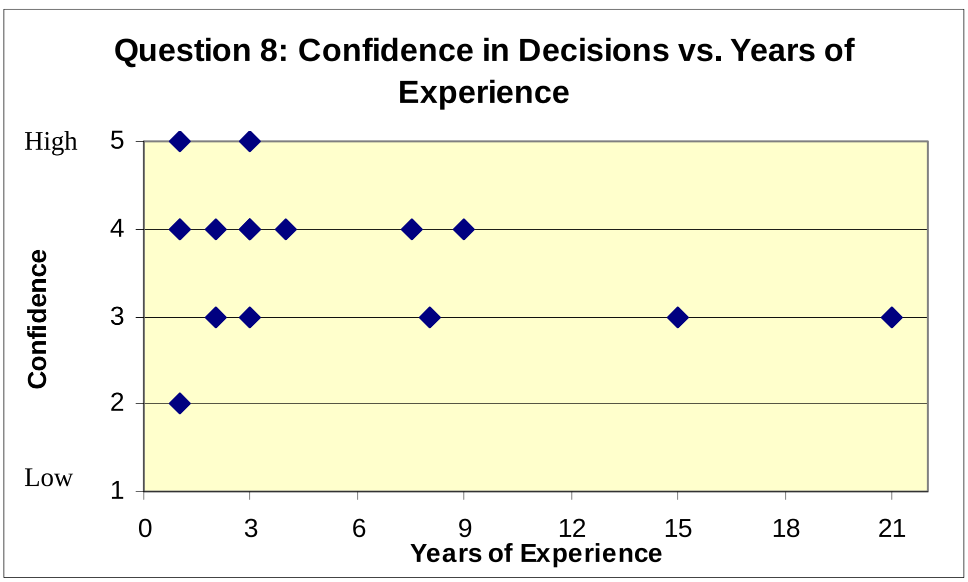


Figure 25. Questionnaire item 8 graph – confidence in decisions

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 users overall to produce better quality COAs?
- Does Weasel help novices more than experts?
- When Weasel exhibits brittle behavior, do some subjects 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?
- Do questionnaire responses provide insight into subject performance or decisions?

The answers to these questions will be used to assess how Weasel should be used and by whom.

6.1 Are evaluators' rankings consistent?

The Spearman Rank Correlation [20] was used to measure the level of agreement between the two evaluators who ranked subject performance in this experiment. Overall, there is a very high level of agreement between the evaluators.

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 identical set. An example is in Scenario 1 where two solutions tied for the third best ranking according to the first evaluator (Figure 13). These two solutions each received a rank value of 3.5 derived from (3 + 4) / 2 = 3.5. The next solution received a rank value of 5, the next rank value in the sequence.

The Spearman Rank Correlation between the two evaluators in Scenario 1 was .904; Scenario 2 was .968; and Scenario 3 was .801. Spearman Rank Correlation calculations are shown in Appendix 3.3.1. The average Spearman Rank Correlation value across the three scenarios was .891. This shows a high to very high level of agreement between the evaluators, on average.

Calculations were also performed on the correlation of quality scores assigned to subject solutions by the evaluators. The correlation values were high, therefore indicating a high level of agreement between the evaluators. The correlation between the two evaluator's quality scores in Scenario 1 was .942; Scenario 2 was .901; and Scenario 3 was .994. The average correlation value across the three scenarios was .946. Therefore, the experimenter felt justified in trusting the evaluator's quality assessments.

Spearman Rank Correlation (r_s) Values
Scenario 1 = .904
Scenario 2 = .968
Scenario 3 = .801
Average Spearman = .891
Quality Score Correlation Values
Scenario 1 = .942
Scenario 2 = .901
Scenario 3 = .994
Average Correlation = .946

Figure 26. Spearman rank correlation values and quality score correlation values

6.2 Does Weasel help users overall to produce better quality COAs?

Yes, the analysis revealed that users produce significantly higher quality ECOAs with the assistance of Weasel than without. The analysis consisted of comparing quality scores for manual ECOAs (Method A) versus ECOAs generated with the assistance of Weasel (Method's B and C). The resulting ANOVA p-value was .018; indicating ECOAs were higher in quality when subjects were assisted by Weasel.

One overall trend was that subjects performed better with Weasel than without across all scenarios. Another trend evident was the computer produced higher quality ECOAs than humans, even when the human was assisted by Weasel. Lastly, as shown in the figure below, the computer performed significantly worse than humans when Weasel exhibited brittle behavior, as in Scenario 3.

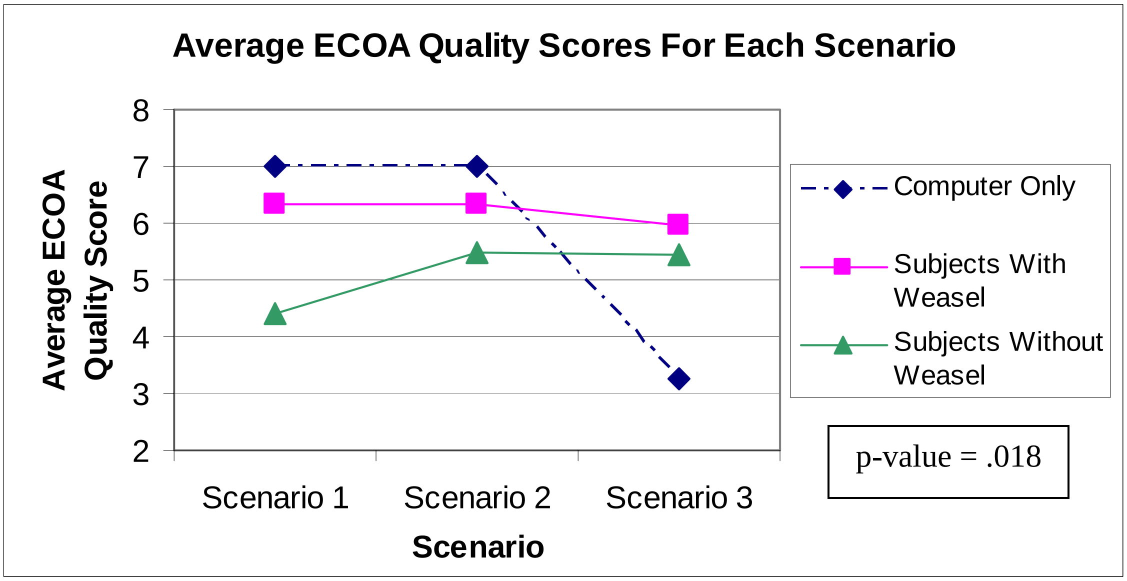


Figure 27. Average ECOA quality for 3 scenarios by type of ECOA generated

6.3 Does Weasel help novices more than experts?

When using automation it is useful to know whether or not the automated tool facilitates improved performance by experts and/or novices. The first step in analyzing performance versus level of experience is to check whether or not subjects really are experts or novices according to their performance before using any type of automation. This was accomplished by the evaluators evaluating the ECOAs generated by each subject before they viewed any automated ECOAs in the respective scenario in which they generated manual ECOAs before viewing computer generated ECOAs, and revisions were not allowed. The graphs in Figures 28 and 30 display ECOA quality scores and subject rankings of expert and novice manual ECOAs prior to any use of automation. The graphs show expert subjects generally produce higher quality ECOAs and receive better performance ranks than novice subjects.

Analysis of Variance (ANOVA) statistics were performed on level of experience versus rank before using automation and experts were ranked significantly higher than novices. The ANOVA returned a p-value of .001. Specific ANOVA data is shown in Appendix 3.3.2. Prior to Weasel assistance, the average rank of experts was 3.80 and the average rank of novices was 11.69. The above information does show subjects were indeed categorized into the appropriate groups prior to completing the scenarios.

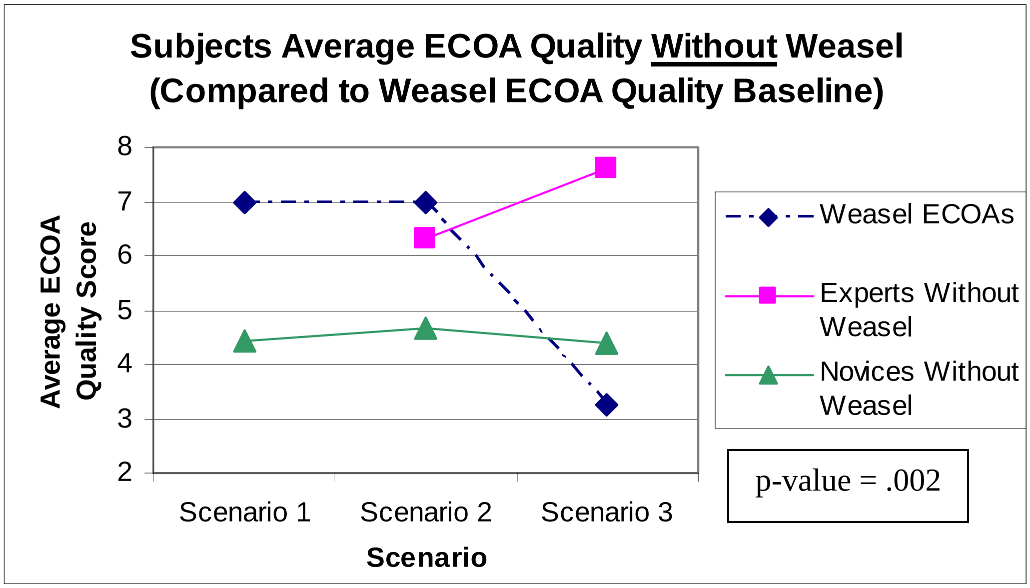


Figure 28. Subject ECOA quality **without** automation use compared to Weasel baseline

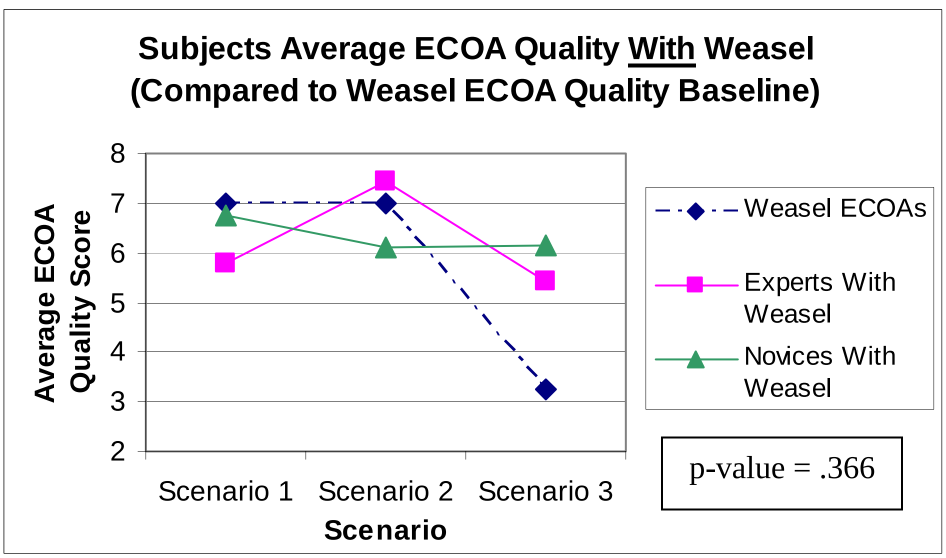


Figure 29. Subject ECOA quality **with** automation use compared to Weasel baseline

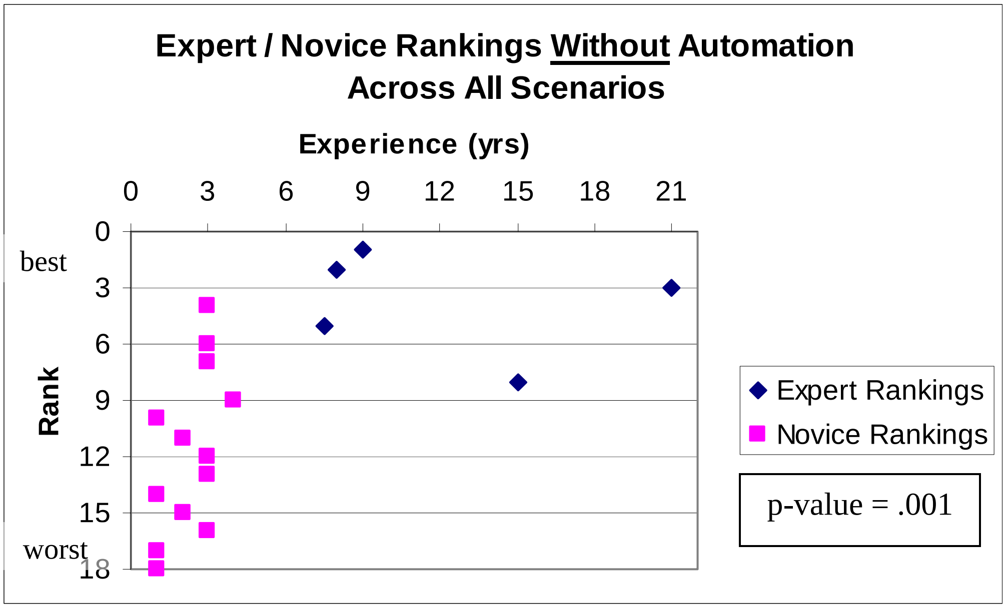


Figure 30. Subject solution rankings **without** automation use (Ranking 1 = best, 18 = worst)

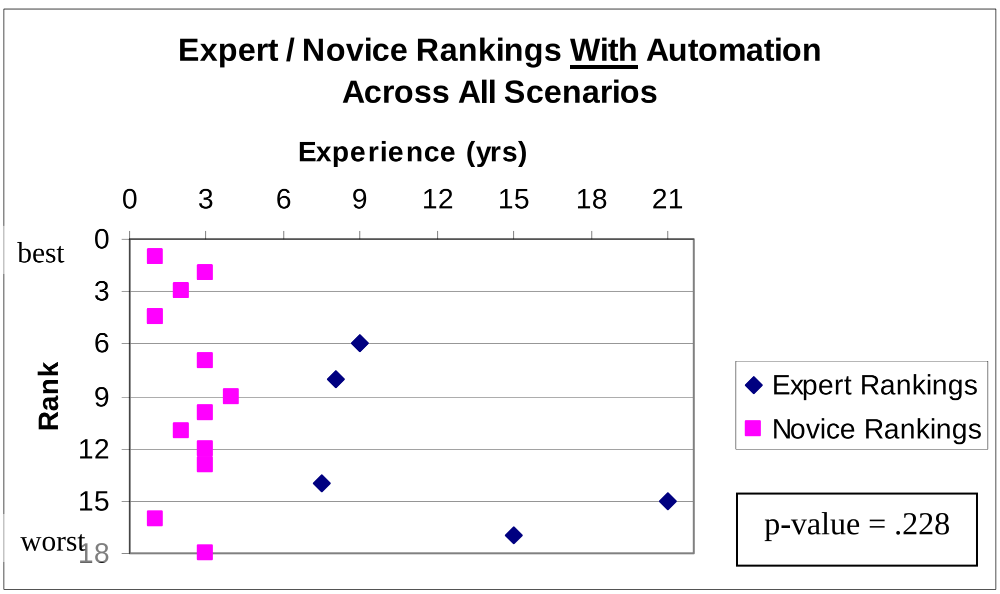


Figure 31. Subject solution rankings **with** automation use (Ranking 1 = best, 18 = worst)

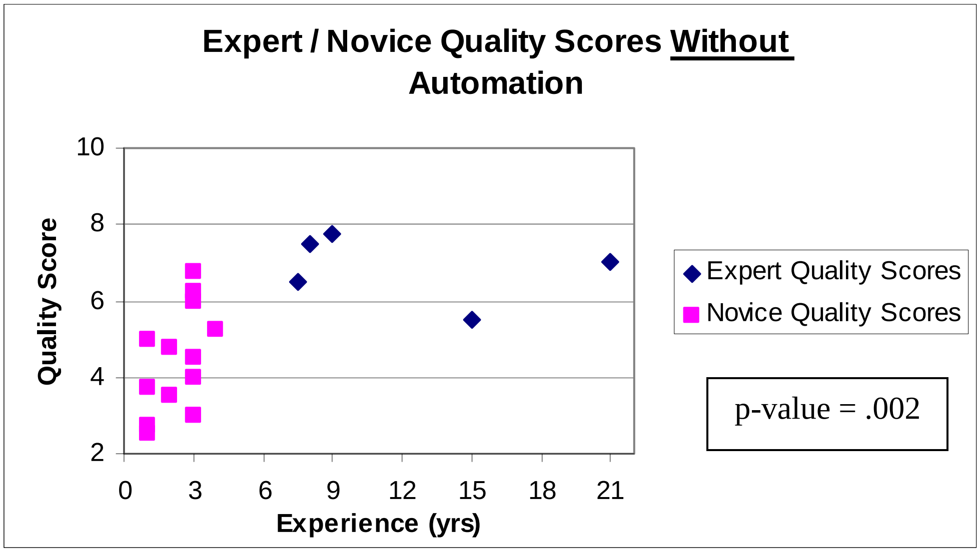


Figure 32. Subject ECOA quality scores **without** automation use (10 = high quality, 0 = low quality)

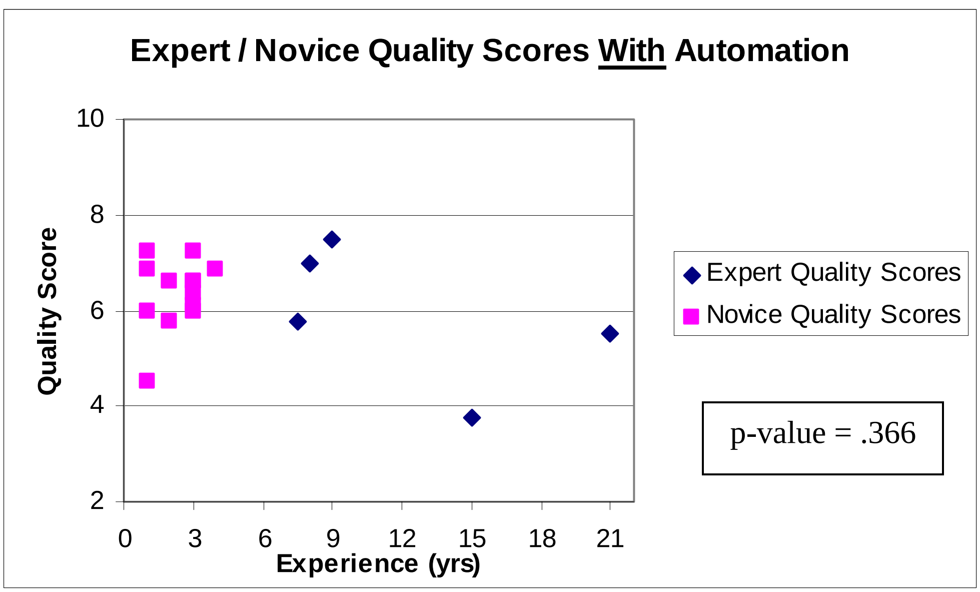


Figure 33. Subject ECOA quality scores **with** automation use (10 = high quality, 0 = low quality)

After using the automation tool, there was no significant difference in ranks between experts and novices. ANOVA statistics (data in Appendix 3.3.2) returned a p-value of .228. In summary, rank evaluation showed experts were significantly better than novices before automation use and there was no difference between experts and novices after automation use. Additionally, the average overall rank across all problems and solution methods of novice subjects was 8.54 while experts were ranked 12 on average.

ANOVA tests were also used to separately analyze the average ranks across all problems for novices and experts when using Weasel versus not using Weasel. According to rankings, there was no significant difference in novice performance with or without Weasel. The resulting p-value was .115. Experts were ranked significantly better without Weasel. The resulting p-value for the expert performance test was .010.

Analysis was performed on ECOA quality scores as well. Before automation use, experts produced ECOAs of significantly higher quality than novices. The resulting p-value was .002. After using Weasel, there was no significant difference between quality scores for experts and novices. The resulting ANOVA p-value was .366. This data is evident in Figure 33 where 17 of 18 subjects had quality scores between 4 and 8.

Novice ECOA quality scores significantly increased with the use of Weasel. The ANOVA p-value was .0001. For expert subjects, there was no significant difference in ECOA quality scores whether or not Weasel was used. The resulting p-value was .251.

Comparison of the data in Figures 28 and 29 show both novice and experts average ECOA quality increased with Weasel assistance relative to the computer generated ECOAs. The exception was expert ECOA quality in Scenario 3. Experts had higher quality ECOAs in Scenario 3 without Weasel.

The usefulness of obtaining quality scores is apparent. Rankings provide performance data for subjects relative to other subjects. Also, since verbal anchors were not used when evaluators assigned quality scores, the rankings provide additional useful information not gained from quality scores alone. Quality scores provide data to compare one subject against another, show a quantifiable difference of how much better or worse subjects are against each other (Figures 32 and 33), and are particularly useful in evaluating data when a group of scores are close together. Another benefit of obtaining ECOA quality score data is the opportunity to see relative performance improvement or degradation for subjects over different scenarios (Figures 28, 29, and 34). Lastly, quality scores of subject performance allow comparison against automated ECOA quality (Figures 28 and 29).

In summary, the data shows Weasel helped both subject groups on average, but helped novices significantly more in terms of generating higher quality ECOAs.

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

Yes, 5 of 18 subjects picked only the Weasel ECOA set in the scenario in which Weasel exhibited brittle behavior. If subjects chose only Weasel's ECOA set, which all showed the enemy leaving one AA uncovered (shown on page 73), the decision was of poor quality. Evaluators felt it was unrealistic to assume the enemy would always leave one AA open. Interestingly, subjects who chose a combination of manual ECOAs (which covered all AAs) and computer generated ECOAs (which left one AA open) received the highest quality scores on average (higher than manual alone or computer alone). The reason being that evaluators felt the combined set covered more potential enemy situations.

Three of the five subjects who chose only Weasel's set were experts (2 novices), and one of the 5 viewed the computer generated ECOAs first (method C). Manual ECOAs generated by subjects before viewing Weasel's ECOA set were analyzed to see if any subjects exhibited the same brittle behavior as the computer (leaving an AA open on all ECOAs). Of 12 subjects who generated manual ECOAs first (Method A & B in scenario 3), only one subject, a novice, left an AA open on all manual ECOAs. The question to ask is whether showing the flawed computer set is inducing some behavior (choosing the flawed solutions) that subjects normally would not choose. Although there is no statistical difference between 1 of 12 (8.33%) subjects producing the same flawed ECOAs as 5 of 18 (27.78%) subjects who choose flawed ECOAs, there may be a practical indication that showing the flawed Weasel set did lead to those ECOAs being chosen more than they would have otherwise.

	Manual Set	**Computer Set**	Manual & Computer Set
Scenario 3 Subject Choices	9/18	**5/18**	4/18

Figure 34. Scenario 3 Subject Choices

6.5 Does ECOA quality decline when Weasel exhibits brittle behavior?

No, analysis showed that the average quality of ECOAs increased when Weasel was used in all three scenarios, including the brittle scenario. Weasel's brittle behavior in Scenario 3 did not significantly degrade user performance. The average quality of all subject's ECOAs were calculated for each of the three scenarios. Comparisons were made between ECOAs generated without Weasel assistance (Method A) and those generated with Weasel assistance (Method's B and C). The data also shows ECOA quality increased in all three scenarios with the use of Weasel. ECOA quality significantly increased with Weasel in Scenario 1 (p-value = .03), and there was no significant difference between the averages in Scenarios 2 (p-value = .31) or Scenario 3 (p-value = .51).

	Scenario 1	Scenario 2	Scenario 3
ECOA quality without Weasel	4.42	5.5	5.46
ECOA quality with Weasel	6.35	6.33	5.96

Figure 35. Average ECOA quality scores with and without Weasel

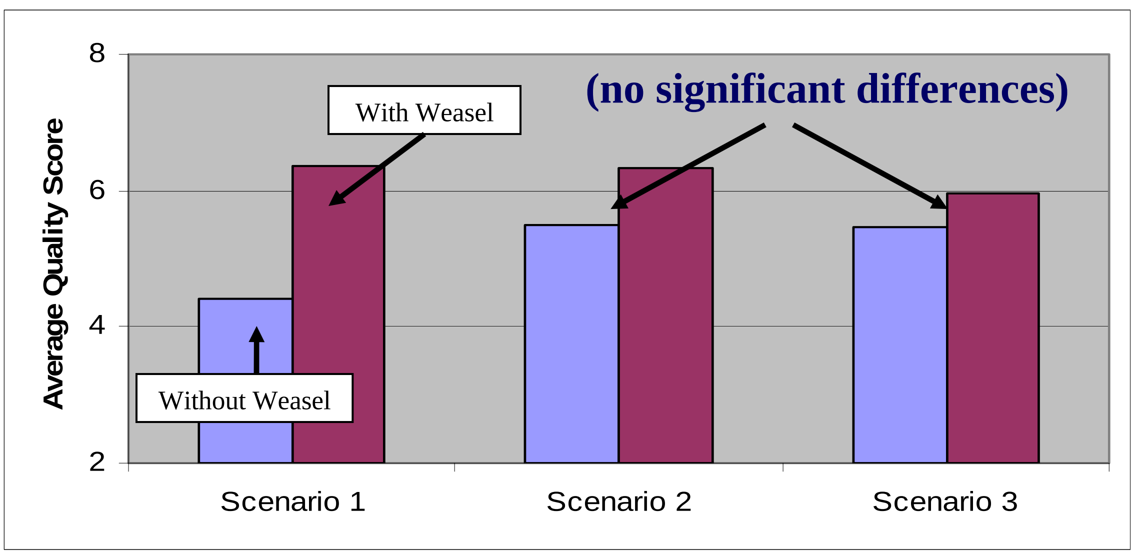


Figure 36. Graph of average ECOA quality scores with and without Weasel

6.6 Does presentation order increase preference toward computer solutions?

There is no significant difference in subject choices between viewing automated ECOAs first and manually generating ECOAs first. 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 .37 which indicates a 37% chance that the results are due to random occurrence. The ANOVA data can be seen in Appendix 3.3.2.

For all scenarios in which subjects viewed the computer generated ECOAs first, only 4 of 18 (22.2%) subjects chose the automated ECOA set as their ideal set. Eight subjects (44.4%) chose their own manual ECOA set as ideal and six (33.3%) chose a mix of both manual and automated ECOAs.

When subjects created manual ECOAs first using problem solving method A, manual ECOA sets were chosen in 16 of 18 decisions (88.9%), automated sets two times (11.1%), and a mix of manual and automated ECOAs were chosen in zero instances.

p-value = .370		**ECOA Generation Method**	
		View Automated First (Method C)	Generate Manual First (Methods A)
Subject Decision	Manual Set	44.4%	88.9%
	Automated Set	22.2%	11.1%
	Mix of Manual & Automated	33.3%	0.0%

Figure 37. Subject decision frequency for ECOA generation method C vs. A

6.7 Does order of presentation impact performance?

Analysis showed that subjects who viewed automated ECOAs first performed better than those that generated manual ECOAs first. The resulting p-value was .018.

To answer this question, analysis was performed comparing subject 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. The average ECOA quality for Method B vs. Method C for each scenario is shown below. Overall, the data indicates viewing automated ECOAs first did not influence subject decisions (section 6.6) but did lead to improved performance.

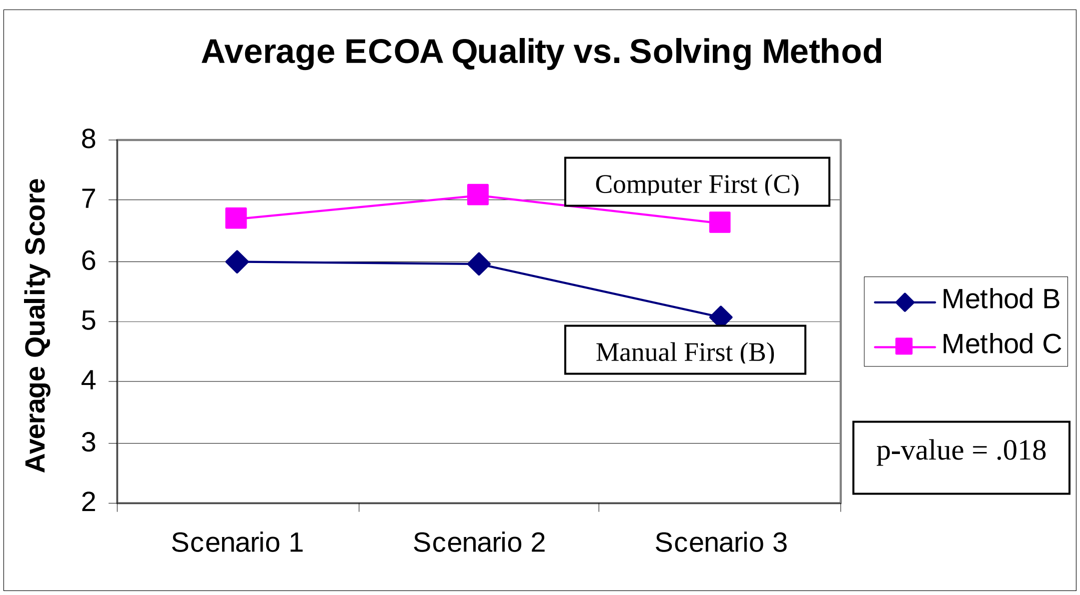


Figure 38. Average ECOA quality vs. solving method

6.8 Do questionnaire responses provide insight into subject performance or decisions?

Responses to relevant questions can often provide useful insight into subject behavior and performance. Analysis was conducted on various questions asked of subjects upon completion of their experimental trials.

Subjects rated their general level of trust in computers in question 1. There was no significant difference in trust in computers between novices and experts as ANOVA analysis gave a p-value of .179. On average novices generally had less trust in computers and this data is supported by subject responses to question 9 which asked whether a human or computer would make a better decision in a situation with multiple variables and potential risk. One of 13 novice subjects (7.69%) selected the computer option to question 9 while two of five experts (40%) selected the computer option. The responses of all subjects on questions 1 and 9 can be seen in Figure 39.

Questionnaire Item #1 – Novice and Expert General Trust in Computers
Rating Scale: 1 (Low Trust) to 5 (High Trust)
and
Item #9 – Better Decision Involving Variables and Risk
Options: Human or Computer

Novice Subject #	**#1 Response**	**#9 Response**	**Expert Subject #**	**#1 Response**	**#9 Response**
1	3	human	11	4	human
2	4	human	13	5	computer
3	5	computer	15	3	human
4	4	human	16	4	human
5	4	human	17	5	computer
6	4	human			
7	4	human			
8	3	human			
9	4	human			
10	3	human	Expert Average On Question 1= 4.20		
12	4	human			
14	3	human			
18	3	human			
Novice Average On Question 1= 3.69					

Figure 39. Novice and expert responses to questionnaire items 1 and 9

The confidence a person has in their decisions and the tools available to them in decision making may play a role in the types of decisions they make. Question 6 asked subjects to rate their confidence in identifying all important ECOAs themselves while question 7 asked subjects to rate their confidence that the computer identified the important ECOAs. ANOVA tests of novice and expert user's confidence in identifying ECOAs returned a p-value of .310, indicating no significant difference in self-confidence between the groups. However, there was a statistically significant difference between the groups in their confidence in the computer identifying important ECOAs. Expert subjects were significantly more confident in computer identification of ECOAs than novice subjects with a p-value for the corresponding ANOVA equal to .043.

Subjects also rated their confidence in the decisions made when choosing between automated and personal ECOAs. Although the average novice rating was higher (3.89) than the average expert rating (3.56) on this question, there was no significant difference between the two sets of confidence ratings. The resulting p-value was .272.

6.9 Overall interpretation of the results

The following are main results from this research:

- Novice performance significantly improved with Weasel,
- Expert performance was not significantly different with Weasel,
- When Weasel exhibited brittle behavior, its use did not significantly change user's average ECOA quality,
- In scenario 3 (where Weasel exhibited brittleness), subjects chose brittle solution sets more often (28% of the time) after seeing Weasel's solution set than before seeing it (where 8% exhibited brittle behavior)
- Presentation order (manual first vs. computer generated ECOAs first) did not influence user preference towards choosing only the computer ECOA set,
- Viewing computer generated ECOAs first significantly increased average ECOA quality.

Useful survey results were more experts than novices felt that a computer would make a better decision than a human in a situation involving many variables and potential risk. Additionally, experts had significantly greater confidence than novices that Weasel would identify all important ECOAs. Surveys also showed there were no significant differences between experts and novices in there overall trust in computers, self-confidence in identifying all important ECOAs, and self-confidence in the decisions made. Both experience groups had high confidence in their own abilities.

Collectively, these results imply that Weasel may be beneficial to use with novice military personnel such as officer trainees in a classroom or lab setting. However, since all DSSs, no matter how good they are, will occasionally exhibit brittle behavior, one should use extreme caution in considering how to use the tool in actual military operations.

Chapter 7 Future Work

In this experiment focus was not placed on issues such as time constraint pressure, group dynamics, and human computer interface interaction. Therefore, there are considerations for future research to be conducted.

The author provided paper copies of ECOAs generated by Weasel to subjects. The automated function to display ECOAs on the map screen was not functioning in the DSS at the time of the experiment. Having subjects view the computer generated ECOAs on the computer itself may be more beneficial for users.

Time constraints were not a consideration in this experiment. The 30-minute time limit in each of the three problem solving scenarios was never reached. 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 this study subjects solved the three scenario problems by themselves. Research analyzing group dynamics when making decisions and using automation tools may be very useful considering the predominance of technology and advanced automated systems available to today's military forces. The decision a person makes when working alone may be quite different than one made when working with another person or an entire group of decision makers. Personalities can be strong in some military members and various types of people may dominate group interaction while others may take a more subtle, quiet role.

Command structure may also influence interaction among group members. The chain of command prevalent in the military plays an important role among members. Decisions may differ when working with subordinates, commanders, or personnel of the same rank.

This experiment strictly utilized the ECOA generator function of the automation tool. There is valuable information to be gained by analyzing the Friendly Course of Action (FCOA) generator as well. The objectives of a military force differ whether they are dealing with enemy or friendly forces. Balancing the two course of action tools adds increased complexity to an already complex situation but also provides useful capabilities to the decision maker(s).

Lastly, additional research analyzing brittle behavior of Weasel would be beneficial. The brittleness Weasel displayed in scenario 3 of this study may have been quite obvious to some subjects. More subtle brittle behavior may impact behavior and performance differently than that seen in this study. 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.

Chapter 8 Conclusions and Recommendations

Based on analysis results and subject feedback there is knowledge to be gained from this experiment and a great deal of potential in the automated tool used. The Weasel ECOA generator is a useful aid that is quite powerful in its capabilities.

The first important result was users overall produced higher quality ECOAs with Weasel assistance than without. The user, working with the automation, generated the best solutions. Additionally, novice ECOA quality significantly improved with the use of Weasel, while expert quality did not change. Therefore, using Weasel with novice military members such as officer trainees (i.e. Academy and or ROTC cadets) would provide beneficial experience in decision making and using decision support tools. Also, despite the results showing expert ECOA quality as the same with or without Weasel, the author feels the tool would be useful in training active duty personnel for battlefield decision making and strategies. This is due to the subjective responses of all subjects, novices and experts alike, that Weasel provided great practice for all experience levels in generating military strategies and was a valuable tool for current and future military personnel. Although time constraints and manual ECOA quantity data were not analyzed for this study, future research may shed light on potential usefulness of experts using Weasel with respect to time savings.

Subjects who viewed the computer generated ECOAs first performed significantly better than those who did not. Based on this result, when using Weasel the author recommends showing computer generated solutions to users first, before any manual COAs are generated.

On a methodological note, gathering ECOA quality scores in addition to performance rankings was quite useful because the two forms of data together complement each other's drawbacks. The drawback to quality scores is that it's difficult to provide verbal anchors when the quality of all solutions being judged is unknown a priori. Without anchors, scores from different judges may be difficult to compare. However, rank data can be compared without verbal anchors. The drawback of rank data is if all solutions judged are very similar in quality, the ranks may lose meaning. However, quality score data may provide a way to identify situations where quality of solutions is similar. The combination of rank and quality data can often avoid the pitfalls of either type of data source alone.

The author also recommends further study on Weasel and the other automation tools that accompany the Intel Tool Kit. Brittleness and time constraints are two specific areas where additional research is needed. Analyzing factors such as group dynamics, brittle behavior, and time constraints will provide further insight into the Intel Tool Kit's influence on user behavior and performance.

In areas as broad as decision making and automation, there is much research to be done in order to fully understand the behavior of human decision makers. This is especially true when the domain involved is as complex as a military environment often is. This research provided important results to enhance understanding of decision support system user performance and behavior. The results and recommendations of this study provide evidence of Weasel's potential to assist users in a military domain.

References

[1] Alreck, Pamela L. and Robert B. Settle. The Survey Research Handbook. 2nd Ed., 1995: 127.

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

[3] Broadbent, D.E. and Margaret Gregory. “Psychological Refractory Period and the Length of Time Required to Make a Decision.” Proceedings of the Royal Society of London. Series B, Biological Sciences. Vol. 168, No. 1011, 1967: 181-193.

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

[5] Cowan, Thomas A. “Decision Theory in Law, Science, and Technology.” Science. Vol. 140, No. 3571, 1963: 1065-1075.

[6] Dzindolet, Mary T., S.A. Peterson, R.A. Pomranky, L.G. Pierce, and H.P. Beck. “The Role of Trust in Automation Reliance.” International Journal of Human-Computer Studies. Vol. 58, 2003: 697-718.

[7] Lee, John D. and Neville Moray. “Trust, Control Strategies, and Allocation of Function in Human-Machine Systems.” Ergonomics. Vol. 35, 1992: 1243-1270.

[8] Lee, John D. and Neville Moray. “Trust, Self-confidence, and Operators’ Adaptation to Automation.” International Journal of Human-Computer Studies. Vol. 40, 1994: 153-184.

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

[10] Montgomery, Douglas C. Design and Analysis of Experiments. Wiley and Sons, 1991: 176-194.

[11] Parasuraman, Raja and Christopher A. Miller. "Trust and Etiquette in High-Criticality Automated Systems." Communications of the ACM. Vol. 47, No. 4, 2004: 51-55.

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

[13] Perrin, Bruce M., B.J. Barnett, L. Walrath, and J.D. Grossman. "Information Order and Outcome Framing: An Assessment of Judgment Bias in a Naturalistic Decision-Making Context." Human Factors. Vol. 43, No. 2, 2001: 227-234.

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

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

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

[17] Smith, Philip J., C. Elaine McCoy, and Charles 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. Vol. 27, No. 3, 1997: 360-371.

[18] Swets, John A. "The Relative Operating Characteristic in Psychology." Science. Vol. 182, No. 4116, 1973: 990-1000.

[19] Vicente, Kim J. Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work. Lawrence Erlbaum, 1999: 46.

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

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

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

Appendices

Appendix 1

Appendix 1.1 Institutional Review Board Approval Letter

Appendix 1.1 contains the approval letter from the University of Minnesota Institutional Review Board (IRB). The approval letter was obtained by the researcher prior to any experimental trials being conducted.

UNIVERSITY OF MINNESOTA

Research Subjects' Protection Programs

Institutional Review Board: Human Subjects Committee (IRB)
Institutional Animal Care and Use Committee (IACUC)

Mayo Mail Code 820
D-528 Mayo Memorial Building
420 Delaware Street S.E.
Minneapolis, MN 55455

612-626-5654
Fax: 612-626-6061
irb@umn.edu
iacuc@umn.edu
www.irb.umn.edu
www.iacuc.umn.edu

March 11, 2004

Adam D. Larson
170 Wedgewood Dr
Mahtomedi MN 55115

Re: "Simulation Decision Making and Trust in Automation"
Human Subjects Code Number: **0403E57169**

Dear Mr. Larson:

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.

The code 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.

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

The IRB wishes you success with this research.

Sincerely,

Cynthia McGill, CIP
Executive Assistant
CLM/aer
CC: Caroline Hayes, Caroline Hayes

Appendix 1.2 Subject Consent Form

Appendix 1.2 contains the consent form that was presented to subjects immediately upon their arrival to the decision support system (DSS) laboratory where each experimental trial was conducted. The DSS lab is in room L121 in the Mechanical Engineering building at the University of Minnesota. Each subject read and signed two copies of the consent form prior to beginning their experimental trial. One copy of the consent form was provided to each subject.

CONSENT FORM
for
Simulation Decision Making and Trust in Automation Research

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: 1Lt Adam Larson, Univ. of Minnesota Department of Industrial 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 3 total hours of your time to be trained in the simulation program and perform the necessary experimental trials. You will sit at a computer workstation 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 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:

You will receive payment via monetary compensation for your time. Compensation will be paid at a rate of $10 per hour, which will be figured in half-hour increments. Payment will be made upon completion of the experiment and will be paid directly to you.

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 researchers conducting this study are: Adam Larson, graduate student, U of MN, and Caroline Hayes (Adam Larson's advisory), Professor, 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 612-624-9850, or via email at lars1770@umn.edu or hayes@me.umn.edu (phone number: 612-626-8390).

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

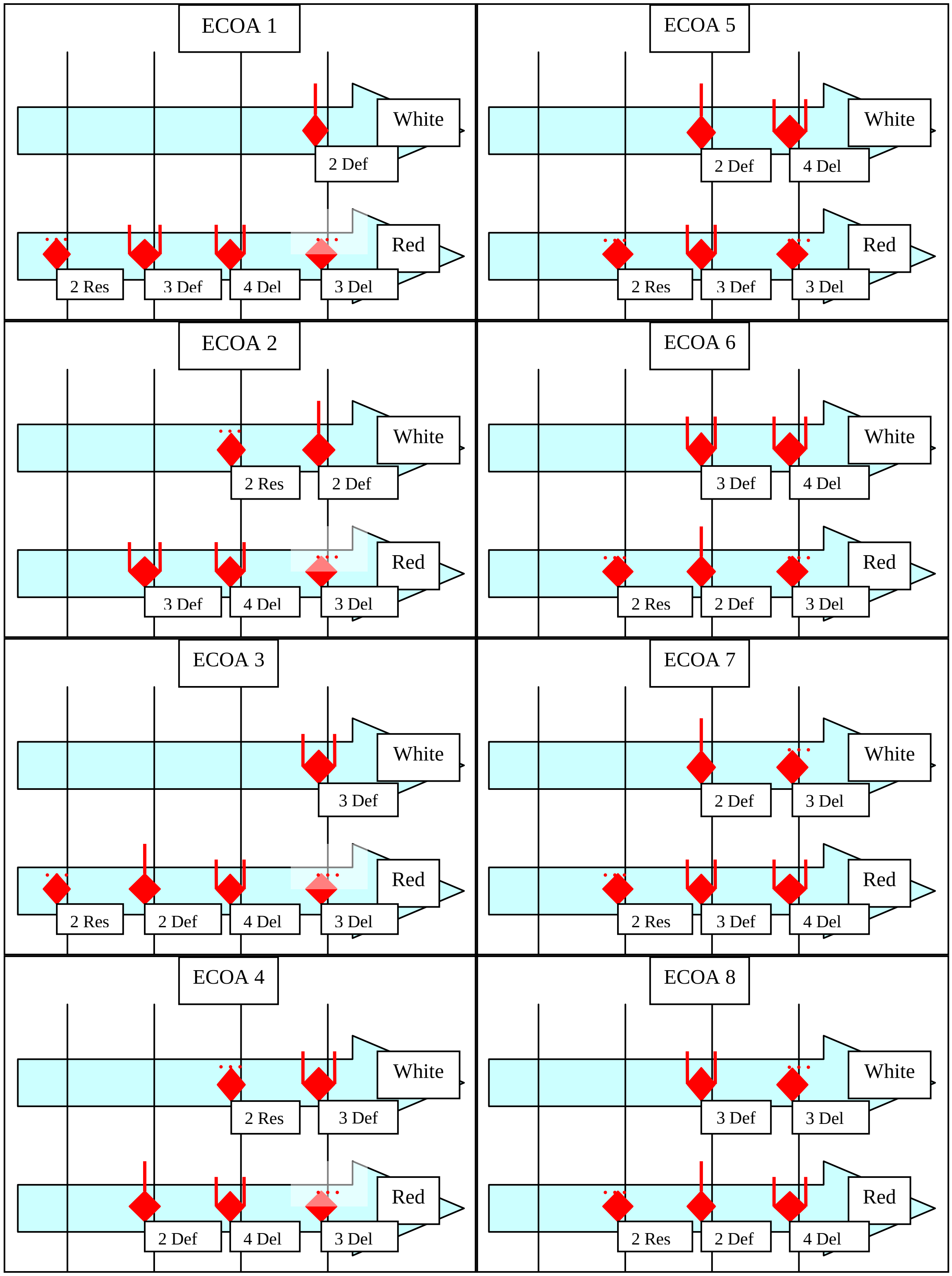
Appendix 2.1 Scenario 1 Description and Weasel ECOAs

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

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.


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 subjects in scenario 2 followed by the 2 computer generated ECOAs from Weasel.

<u>Scenario 2</u>

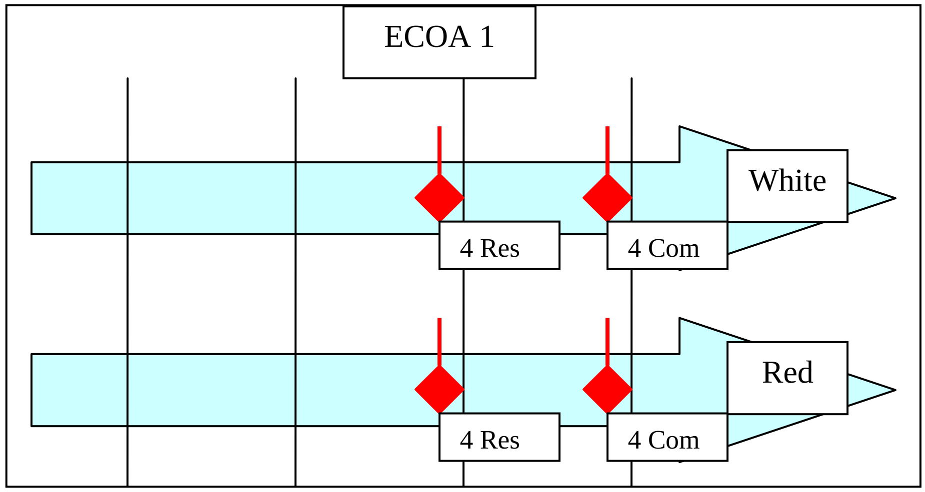
You've received intelligence from computer surveillance and analysis that enemy forces are on the move, heading towards your allied forces. To best prepare your defense, you must identify any possible courses of action the enemy may take in attacking your forces. You are confident in the following enemy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

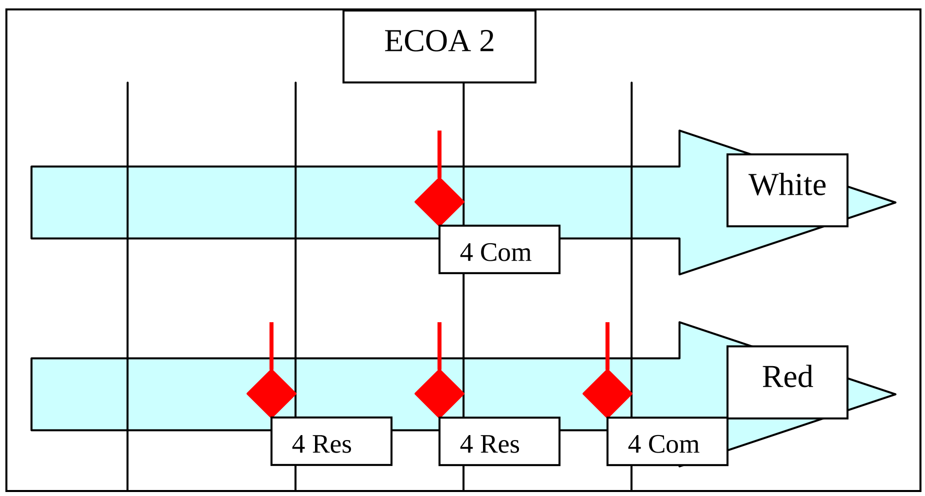
- ❖ Intention is to attack from the west along 2 Avenues of Approach (AA_white is north of AA_red).

- ❖ 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.


ECOA 1
White
4 Res
4 Com
Red
4 Res
4 Com


ECOA 2
White
4 Com
Red
4 Res
4 Res
4 Com

Appendix 2.3 Scenario 3 Description and Weasel ECOAs

Appendix 2.3 is the situational description given to subjects in scenario 3 followed by the 8 computer generated ECOAs from Weasel.

<u>Scenario 3</u>

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).

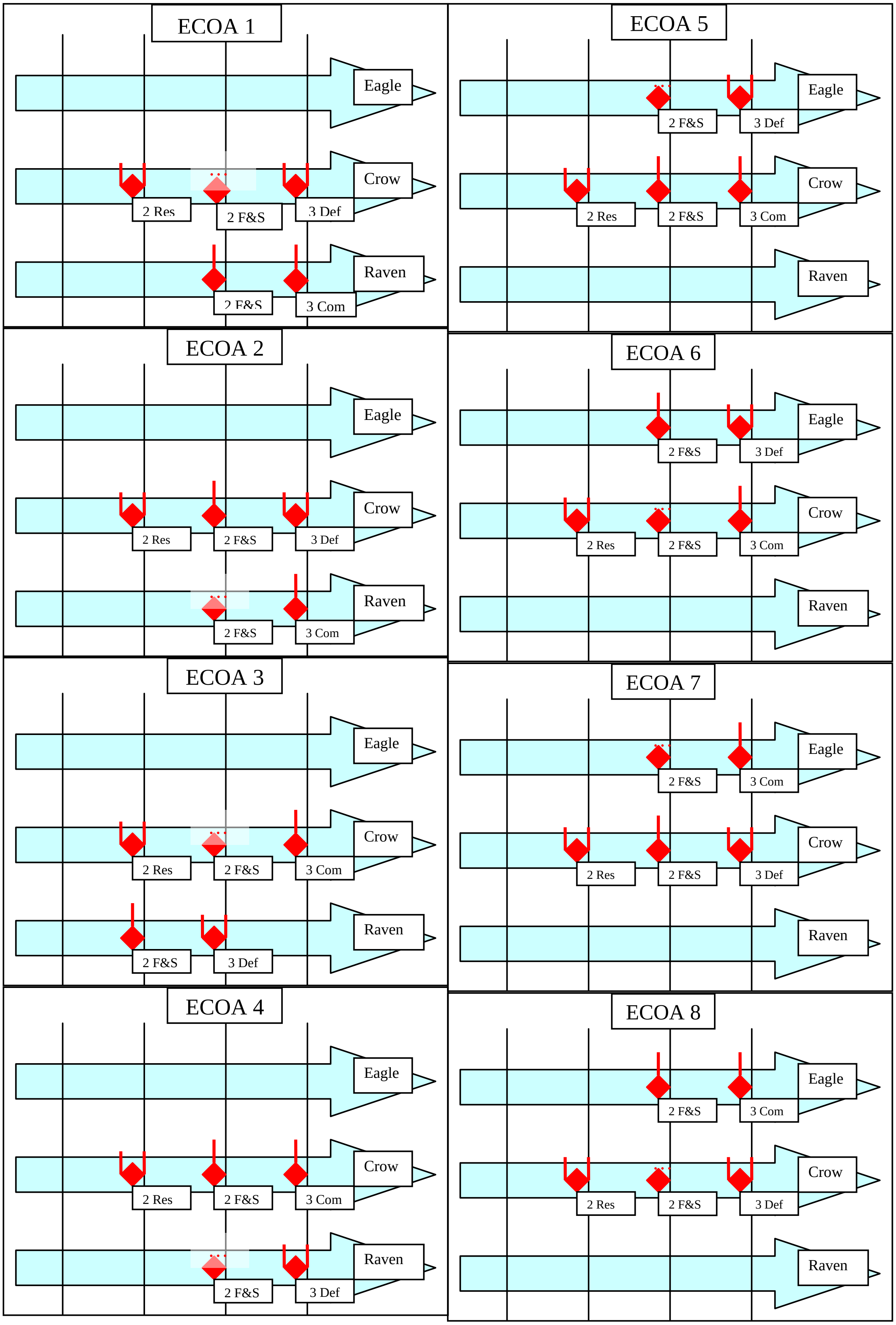
- ❖ Enemy forces consist of 2 companies, 2 battalions, and 1 platoon. Details on each unit follows:

    - ➢ 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
      - ▪ <u>Note:</u> 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
Eagle
Crow
2 Res
2 F&S
3 Def
Raven
2 F&S
3 Com
ECOA 2
Eagle
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
Eagle
2 F&S
3 Com
Crow
2 Res
2 F&S
3 Def
Raven
ECOA 8
Eagle
2 F&S
3 Com
Crow
2 Res
2 F&S
3 Def
Raven

Appendix 2.4 Subject Questionnaire

Appendix 2.4 contains the questionnaire given to each subject after the 3 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

Appendix 2.5 Data Recording Form

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

Name:			**Subject #:**	
Date:				
Military Service:	Air Force	Army		
Subject Type:	Expert	Novice	**If Expert, Yrs. Experience:**	
Simulation Training Time (min):			**If Novice, Yrs in ROTC:**	
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:			
Scenario 3:				
	# Personal ECOAs generated:			
Time:	# Personal ECOAs revised:			
	ECOA set chosen: Personal	Automated		Both
	Reason for decision:			
	Verbal comments during scenario 3:			

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 subject prior to beginning familiarization training on the DSS tool.

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

Thesis Research by: 1Lt 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.

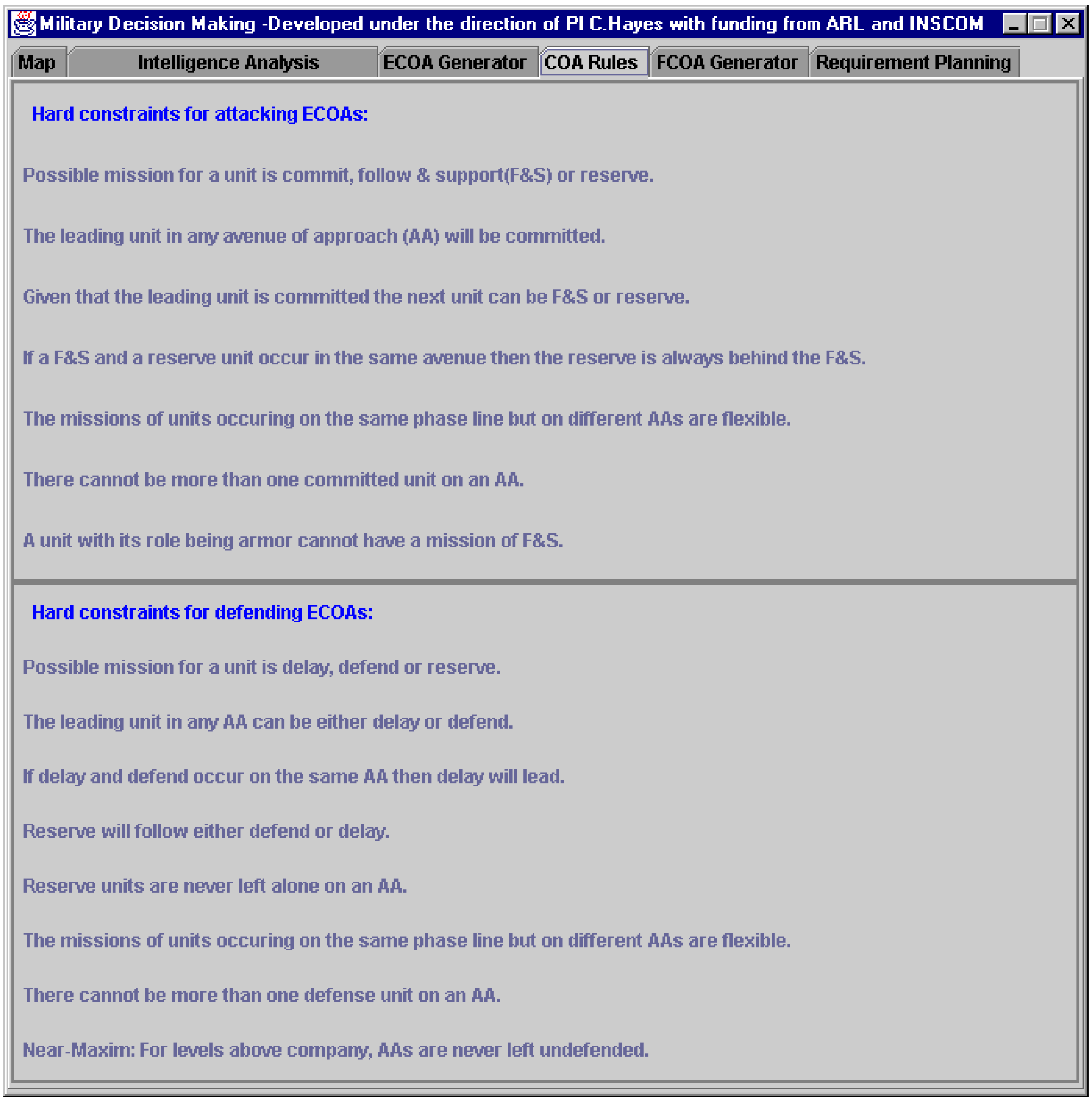
3. The task of subjects will be to evaluate intelligence information, formulate potential enemy courses of action, and analyze courses of action generated by the computer. Considering the given circumstances outlined in the scenario and all relevant information, a decision will then be made to choose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

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

Respectfully,
Adam Larson, 1Lt, USAF

Appendix 2.7 Decision Support System (DSS) Constraints

Appendix 2.7 shows the one page screen capture that was provided to each subject. These constraints were discussed during familiarization training and subjects were instructed to abide by them to the best of their abilities when generating and analyzing ECOAs. Subjects were provided as much time as they needed to study and understand these constraints before beginning the scenario problems.



Appendix 3

Appendix 3.1 Subject ECOA Rankings Prior to Automation Use

This appendix displays subject rankings for ECOAs manually generated prior to the subject viewing automated ECOAs. Included is the scenario problem in which the subject manually generated ECOAs first and was not allowed to revise manual ECOAs.

Subject	Experience Level	Evaluator Rank Prior to Use of Automation	Scenarios Evaluated
1	Novice	18	1
2	Novice	11	1
3	Novice	4	1
4	Novice	7	1
5	Novice	15	1
6	Novice	6	2
7	Novice	10	2
8	Novice	14	3
9	Novice	17	2
10	Novice	12	3
12	Novice	9	3
14	Novice	13	3
18	Novice	16	1
11	Expert	5	2
13	Expert	2	3
15	Expert	3	2
16	Expert	1	3
17	Expert	8	2

Appendix 3.2 Data on Subject Time to Complete Scenario Problems

Subject #	Experience (yrs)	Problem Solving Method	Experience Level	Scenario 1 Decision	Scenario 1 Completion Time (min)	Scenario 2 Decision	Scenario 2 Completion Time (min)	Scenario 3 Decision	Scenario 3 Completion Time (min)
1	1	2B/1A/3C	Novice	Manual	11	Manual	8	Manual	12
7	1	3C/2A/1B	Novice	Both	16	Manual	10	Both	20
8	1	3A/1C/2B	Novice	Automated	24	Manual	16	Manual	28.5
9	1	2A/1C/3B	Novice	Automated	20	Manual	11	Automated	19
2	2	1A/3C/2B	Novice	Manual	12	Manual	6	Both	23
5	2	3B/1A/2C	Novice	Manual	11.5	Manual	3	Both	23
4	3	2C/3B/1A	Novice	Manual	9	Manual	6	Automated	16
6	3	1B/3C/2A	Novice	Both	28.5	Manual	7	Manual	16
14	3	3A/2C/1B	Novice	Both	28	Manual	11	Manual	24
18	3	1A/2C/3B	Novice	Manual	27	Both	4	Both	25
3	3	3C/2B/1A	Novice	Manual	6	Both	4	Manual	15
10	3	1C/2B/3A	Novice	Both	14	Manual	5	Manual	13
12	4	2C/3A/1B	Novice	Both	20	Both	8	Manual	17
11	7.5	1C/3B/2A	Expert	Manual	22	Manual	6	Manual	15
13	8	1B/2C/3A	Expert	Automated	7.5	Automated	2	Manual	9
16	9	2B/3A/1C	Expert	Both	11	Both	12	Automated	19
17	15	2A/1B/3C	Expert	Manual	13	Automated	7	Automated	8
15	21	3B/2A/1C	Expert	Manual	15	Manual	9	Automated	19

Note: Example Problem Solving Method 2B/1A/3C

- Solve Scenario 2 first with Method B (manual first, then automated, revisions)
- Solve Scenario 1 second with Method A (manual first, then automated)
- Solve Scenario 3 third with Method C (automated first, then manual)

Appendix 3.3 Statistical Analysis

Appendix 3.3.1 Spearman Rank Correlation calculations

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

SPEARMAN RANK CORRELATION – SCENARIO 1					
Subject	**Experience Level**	**Evaluator 1 Rank**	**Evaluator 2 Rank**	**Rank Difference**	**Difference Squared**
1	Novice	12	13.5	1.5	2.25
2	Novice	1	1	0	0
3	Novice	13.5	11	2.5	6.25
4	Novice	8	9.5	1.5	2.25
5	Novice	8	3	5	25
6	Novice	8	6.5	1.5	2.25
7	Novice	3.5	2	1.5	2.25
8	Novice	11	12	1	1
9	Novice	3.5	4	.5	0.25
10	Novice	13.5	13.5	0	0
12	Novice	5	6.5	1.5	2.25
14	Novice	16	17	1	1
18	Novice	15	15	0	0
11	Expert	17	16	1	1
13	Expert	8	5	3	9
15	Expert	2	8	6	36
16	Expert	8	9.5	1.5	2.25
17	Expert	18	18	0	0
				SUM =	**93**
				Scenario 1 r_s =	**.9040**

SPEARMAN RANK CORRELATION – SCENARIO 2					
Subject	**Experience Level**	**Evaluator 1 Rank**	**Evaluator 2 Rank**	**Rank Difference**	**Difference Squared**
1	Novice	17.5	18	.5	0.25
2	Novice	16	15	1	1
3	Novice	11.5	10.5	1	1
4	Novice	11.5	12	.5	0.25
5	Novice	11.5	10.5	1	1
6	Novice	5	7.5	2.5	6.25
7	Novice	9	7.5	1.5	2.25
8	Novice	7	5	2	4
9	Novice	1	1	0	0
10	Novice	15	16	1	1
12	Novice	11.5	14	2.5	6.25
14	Novice	3	2	1	1
18	Novice	3	4	1	1
11	Expert	14	13	1	1
13	Expert	7	9	2	4
15	Expert	17.5	17	.5	0.25
16	Expert	3	3	0	0
17	Expert	7	6	1	1
				SUM =	**31.5**
				Scenario 2 r_s =	**.9675**

SPEARMAN RANK CORRELATION – SCENARIO 3					
Subject	**Experience Level**	**Evaluator 1 Rank**	**Evaluator 2 Rank**	**Rank Difference**	**Difference Squared**
1	Novice	7.5	8.5	1	1
2	Novice	1	1	0	0
3	Novice	9	8.5	.5	0.25
4	Novice	6	10	4	16
5	Novice	15	12	3	9
6	Novice	2	5	3	9
7	Novice	3	3	0	0
8	Novice	4.5	4	.5	0.25
9	Novice	17	17	0	0
10	Novice	16	15	1	1
12	Novice	13	6.5	6.5	42.25
14	Novice	7.5	6.5	1	1
18	Novice	13	13	0	0
11	Expert	4.5	2	2.5	6.25
13	Expert	13	11	2	4
15	Expert	18	14	4	16
16	Expert	10.5	16	5.5	30.25
17	Expert	10.5	18	7.5	56.25
				SUM =	**192.5**
				Scenario 3 r_s =	**.8013**

Appendix 3.3.2 ANOVA Calculations

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

ANOVAs for Section 6.2

Does Weasel help users overall to produce better quality ECOAs? YES

ECOA Quality Scores With (Method B/C) Weasel vs. Without (Method A) Weasel						
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	11.25043	1	11.25043403	6.16812	**0.0180976**	4.130017
Within Groups	62.01476	34	1.82396344			

Table above shows significantly higher quality ECOAs with Weasel than without Weasel.

ANOVAs for Section 6.3

Does Weasel help novices more than experts? YES

Rank Before Automation Use vs. Experience Level						
Groups	*Count*	*Sum*	*Average*	*Variance*		
experts	5	19	3.8	7.7		
novice	13	152	11.69231	19.0641		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	224.9308	1	224.9308	13.86486	**0.001848**	4.493998
Within Groups	259.5692	16	16.22308			

Table above shows experts ranked significantly better than novices before automation use.

Rank After Automation Use vs. Experience Level						
Groups	*Count*	*Sum*	*Average*	*Variance*		
experts	5	60	12	22.5		
novices	13	111	8.538462	29.22756		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	43.26923	1	43.26923	1.570818	**0.228095**	4.493998
Within Groups	440.7308	16	27.54567			

Table above shows no significant difference between expert and novice rankings after using the automation tool.

Quality Scores Before Automation Use vs. Experience Level						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Novices	13	58	4.4615385	1.862981		
Experts	5	34.25	6.85	0.8		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	20.6004808	1	20.600481	12.89758	**0.002444**	4.493998
Within Groups	25.5557692	16	1.5972356			

Table above shows experts had significantly higher quality ECOAs before using automation than novices.

Quality Scores After Automation Use vs. Experience Level						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Novices	13	82.875	6.375	0.539063		
Experts	5	29.5	5.9	2.14375		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	0.81475694	1	0.8147569	0.866547	**0.365747**	4.493998
Within Groups	15.04375	16	0.9402344			

Table above shows no significant difference between expert and novice ECOA quality after use of automation.

Novice Ranks With Weasel vs. Without Weasel						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Before Automation	13	152	11.69231	19.0641		
After Automation	13	111	8.538462	29.22756		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	64.65385	1	64.65385	2.67764	**0.114815**	4.259677
Within Groups	579.5	24	24.14583			

Table above shows no significant difference in novice's rankings with or without the use of Weasel.

Experts Ranks Without Weasel vs. With Weasel						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Before Automation	5	19	3.8	7.7		
After Automation	5	60	12	22.5		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	168.1	1	168.1	11.13245	**0.010284**	5.317655
Within Groups	120.8	8	15.1			

Table above shows expert's rankings were significantly better before using Weasel.

Novice Quality Scores With & Without Automation						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Without Automation	13	58	4.4615385	1.862981		
With Automation	13	82.875	6.375	0.539063		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	23.7986779	1	23.798678	19.81536	**0.000168**	4.259677
Within Groups	28.8245192	24	1.2010216			

Table above shows novice user quality scores were significantly better with the use of the automation tool.

Expert Quality Scores With & Without Automation						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Without Automation	5	34.25	6.85	0.8		
With Automation	5	29.5	5.9	2.14375		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	2.25625	1	2.25625	1.532909	**0.250776**	5.317655
Within Groups	11.775	8	1.471875			

Table above shows no significant difference in expert user quality scores with or without use of the automation tool.

ANOVAs for Section 6.6

Does presentation order increase preference toward computer solutions? NO

Decision When Viewing Automated ECOAs First vs. Decision When Generating Manual ECOAs First						
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.694444	1	0.694444	0.841584	**0.365406**	4.130018
Within Groups	28.05556	34	0.825163			

Table above shows no significant difference in decisions made whether subjects viewed automated ECOAs first or manually generated ECOAs first.

ANOVA for Section 6.7

Does order of presentation impact performance? YES

Method B subject rankings vs. Method C subject rankings						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Method B	18	102.25	5.680555	2.851511		
Method C	18	122.5	6.805555	0.849673		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	11.390625	1	11.390625	6.155123	**0.018210**	4.130017
Within Groups	62.920138	34	1.85059232			

Table above shows Method C (viewing automated ECOAs first) subjects rank significantly better than Method B (manually generated ECOAs first) subjects.

ANOVAs for Section 6.8

Question 1 did not have a statistically significant difference between expert and novice subject's trust in computers.

Survey Question 1 - Expert vs. Novice general trust in computers						
Groups	*Count*	*Sum*	*Average*	*Variance*		
novices	13	48	3.692308	0.397436		
experts	5	21	4.2	0.7		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	0.930769	1	0.930769	1.96748	**0.179821**	4.493998
Within Groups	7.569231	16	0.473077			

Question 6 did not have a statistically significant difference between experts and novices self confidence in identifying important ECOAs.

Survey Question 6 – Expert vs. Novice self confidence in identifying ECOAs						
Groups	*Count*	*Sum*	*Average*	*Variance*		
novices	13	43	3.307692	0.897436		
experts	5	14	2.8	0.7		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	0.930769	1	0.930769	1.097506	**0.310383**	4.493998
Within Groups	13.56923	16	0.848077			

Question 7 did have a statistically significant difference between expert and novice confidence in the computer identifying important ECOAs. Experts had significantly greater confidence in the computer than novices.

Survey Question 7 – Expert vs. Novice confidence in computer identifying ECOAs						
Groups	*Count*	*Sum*	*Average*	*Variance*		
novices	13	35	2.692308	0.397436		
experts	5	17	3.4	0.3		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	1.808547	1	1.808547	4.847652	**0.042702**	4.493998
Within Groups	5.969231	16	0.373077			

Question 8 did not have a statistically significant difference between subject groups in the confidence of their decisions.

Survey Question 8 – Expert vs. Novice confidence in decisions						
Groups	*Count*	*Sum*	*Average*	*Variance*		
novices	13	50	3.846154	0.641026		
experts	5	17	3.4	0.3		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	0.718803	1	0.718803	1.293349	**0.272177**	4.493998
Within Groups	8.892308	16	0.555769			