2014-06

A comparison of tactical leader decision making between automated and live counterparts in a virtual environment

Patton, Scott A.

Monterey, California: Naval Postgraduate School

http://hdl.handle.net/10945/42705

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

A COMPARISON OF TACTICAL LEADER DECISION MAKING BETWEEN AUTOMATED AND LIVE COUNTERPARTS IN A VIRTUAL ENVIRONMENT

by

Scott A. Patton

June 2014

Thesis Advisor: Quinn Kennedy
Second Reader: Jonathan Alt

This thesis was performed at the MOVES Institute

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704–0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704–0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2014	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE A COMPARISON OF TACTICAL LEADER DECSISION MAKING BETWEEN AUTOMATED AND LIVE COUNTERPARTS IN A VIRTUAL ENVIRONMENT			**5. FUNDING NUMBERS**
6. AUTHOR(S) Scott A. Patton			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943–5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB protocol number ____NPS.2014.009-IR ____.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release;distribution is unlimited			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

The use of "responsible" autonomous systems may not be far away. Prior to developing or using responsible autonomous systems, it may be important to know if tactical leaders would make different types of decisions with automated systems than they would make with a human live crew. This work attempts to determine if decisions, time to make decisions, and confidence in decisions differ when tactical leaders rely on an autonomous wingman or a live wingman.

Virtual Battlespace Simulation 2 was used to provide the virtual environment in which 30 military personnel completed a simulated mission that entailed five decision points. Participants were randomly assigned to have an autonomous or live wingman.

Decision patterns were compared to a standard based on Army Doctrine for mechanized infantry Bradley sections and subject matter experts. Results indicated no significant group difference in decisions made, time to make decisions, and confidence in decisions. However, significant group differences emerged in the aspects of the wingman that participants trusted most and least. Although most participants indicated that they would not trust autonomous wingmen in real combat, results suggest that participants would revert to doctrinal decisions when faced with an unambiguous situation with an unmanned system with which they had some experience.

14. SUBJECT TERMS Decision Making, Unmanned Systems, Human Robotic Integration			**15. NUMBER OF PAGES** 119
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540–01–280–5500 Standard Form 298 (Rev. 2–89) Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

A COMPARISON OF TACTICAL LEADER DECSISION MAKING BETWEEN AUTOMATED AND LIVE COUNTERPARTS IN A VIRTUAL ENVIRONMENT

Scott A. Patton
Major, United States Army
B.S., United States Military Academy, 2002

Submitted in partial fulfillment of the
requirements for the degree of

**MASTER OF SCIENCE IN
MODELING, VIRTUAL ENVIRONMENTS, AND SIMULATION**

from the

**NAVAL POSTGRADUATE SCHOOL
June 2014**

Author: Scott A. Patton

Approved by: Quinn Kennedy, Ph.D.
Thesis Advisor

Jonathan Alt
Second Reader

Chris Darken, Ph.D.
Chair, MOVES Academic Committee

Peter Denning, Ph.D.
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The use of “responsible” autonomous systems may not be far away. Prior to developing or using responsible autonomous systems, it may be important to know if tactical leaders would make different types of decisions with automated systems than they would make with a human live crew. This work attempts to determine if decisions, time to make decisions, and confidence in decisions differ when tactical leaders rely on an autonomous wingman or a live wingman.

Virtual Battlespace Simulation 2 was used to provide the virtual environment in which 30 military personnel completed a simulated mission that entailed five decision points. Participants were randomly assigned to have an autonomous or live wingman.

Decision patterns were compared to a standard based on Army Doctrine for mechanized infantry Bradley sections and subject matter experts. Results indicated no significant group difference in decisions made, time to make decisions, and confidence in decisions. However, significant group differences emerged in the aspects of the wingman that participants trusted most and least. Although most participants indicated that they would not trust autonomous wingmen in real combat, results suggest that participants would revert to doctrinal decisions when faced with an unambiguous situation with an unmanned system with which they had some experience.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AI	Artificial Intelligence
ARL	Army Research Laboratory
BI	Bohemia Interactive
BDA	Battle damage assessment
BP	Battle position
CIV	Commander's independent viewer
CP	Check point (general followed by number)
DIS	Distributed interactive simulation
FBCB2	Future Battle Command Brigade and Below
HMMWV	High-mobility medium wheeled vehicle
HRI	Human-Robot interaction
HUD	Heads-up-display
IED	Improvised explosive device
ITT	Immersive Training Trainer
RPD	Recognition-primed decision
SD	Standard deviation
SJT	Situational judgment test
UAV	Unmanned aerial vehicles
UGV	Unmanned ground vehicles
VBS2	Virtual Battlespace Simulation 2

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

No human endeavor is ever truly conducted alone. Even marathon runners have a support network of friends and family who encouraged them during training, cheered from the sidelines, or offered aid. Likewise, this project also had a great deal of support.

First and foremost, I would like to thank my thesis advisors. Dr. Quinn Kennedy acted as my Thesis Advisor, evaluating all documents and shaping the experimental design to ensure academic rigor and standards were maintained. LTC Jonathan Alt provided both analytical expertise and military knowledge. MAJ Peter Nesbitt provided a sounding board for ideas and concepts that helped the initial development of the study, and ensured it maintained relevance to the larger Army Studies.

TCM-Gaming provided support for the use of VBS2, providing the software, licensing and an administrative class on the use and function of VBS2; as well as an introduction to scripting techniques and methods. In addition, Robert Munsey and Robert Chavez assisted with the development the graphic user interface display. Their continued support throughout the project was instrumental in achieving the study goals.

Within the MOVES department, Professor Don McGregor assisted with the development of a Websocket Multi-thread Gateway allowing the creation of an external VBS2 map. This map capability allowed for a more realistic and robust scenario while ensuring full screen functionality for conducting eye-tracking. Erik Johnson and his support staff at NPS provided valuable technical insight and advice for both the functionality within VBS2 as well as the use of the LVC Gateway. Without their advice, the physical entity representation could not have occurred.

Lastly, Jesse Huston was probably the most utilized of personnel resources available. Mr. Huston assisted with the collection of data and with the conduct of the experiment. In order to maintain the ruse that a live wingman was working with the participant, Mr. Huston had to play a more up front role in the administering of the study, resolving issues, and maintaining scenario compliance.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. PROBLEM STATEMENT

In 2012, the Department of Defense established a science board task force to review the creation and study of autonomous systems. The board's findings established the need to focus more on the interaction between humans and computers/robots (Defense Science Board, OUSD(AT&L), 2012). Carrying on this theme, the Army Capabilities and Integration Command commissioned a study by RAND Arroyo on the benefits of pursuing unmanned systems (Robotic Systems, Joint Project Office (RS JPO), 2011). While the study concluded that there were many benefits to unmanned ground systems (lower cost, fewer casualties, etc.), it also outlined some of the challenges: acceptance into the military culture and adaptation of tactics, techniques, and procedures that incorporate unmanned ground vehicles (Robotic Systems, Joint Project Office [RS JPO], 2011). The challenge with implementing unmanned systems is ensuring that the Soldiers and leaders at the tactical level will use them, and discovering how or why they will use them. This study attempts to determine current trust levels and review the types of decisions tactical leaders make with unmanned systems. The purpose is to provide evidence of consensus for specific tactics, techniques, and procedures when using unmanned systems, allowing for more refined development and fielding.

B. RESEARCH QUESTIONS AND CORRESPONDING HYPOTHESES

Three research questions guided the study to explore decision making within military tactical environments while using automated ground combat systems.

1. Do a leader's decisions differ when using an automated system versus a live crew as a Bradley Section wingman?

H_0**:** The mean accuracy for each group's decisions will be the same.

H_A**:** The mean accuracy for each group's decisions will be different.

2. Does the amount of time taken to make the decisions differ when the leader has an autonomous wingman versus a live wingman?

H0: The mean time for a decision by each group will be the same.

HA: The mean time for a decision by each group will be different.

3. Is there a difference in the leader's confidence when using an automated wingman versus live?

H_0: The confidence level of each group will be the same.

H_A: The confidence level of each group will be different.

C. EXPLORATORY QUESTION

- What are the benefits and drawbacks to using Virtual Battlespace Simulation 2 (VBS2) as a means for testing future tactical concepts?

D. SCOPE OF THESIS

An Army-approved simulation (Virtual Battlefield Simulation 2) was used to compare vehicle section leader decisions with and without an unmanned ground vehicle (UGV) counterpart (wingman) to that of a live wingman. The study supports U.S. Army interest in UGV development, informs the larger Army Research Office-sponsored investigations into tactical decision making and reinforces VBS2 as both a trainer and a research tool. The level of leader interaction within VBS2 is limited in order to force specific decisions and allow for clearly defined decision paths. All decisions and information were provided to the participant via a heads-up-display (HUD) and on-screen video, providing a first-person perspective. We can assume that a UGV-like display will return similar results to any future design. There were a total of 98 possible paths with two basic types of decisions, movement and tactical, which can affect the path. Movement decisions define whether the wingman or leader (i.e., the participant) will move first in the next action. Tactical decisions determine a main action, such as defend or assault. We refer to the participants' chain of decisions as the decision path. Each decision retains a value that expresses how far away from what the study team expected the decision to be. We determined the value by a review of current doctrine and by subject matter experts (SMEs), so that a total decision path receives a quantifiable value called a path score. The lower the path score, the closer the actions were that would be expected of a live situation. Active and retired U.S. service members from the Naval Postgraduate School (NPS), many of whom have experience with unmanned systems and scenario-specific activities, comprised the sample of participants. Participants were blocked by level of experience. The level of experience was determined using a metric of the participant's demographic data, but was most heavily weighted to those with unmanned systems or mounted ground combat (high, mid, low).

E. BENEFITS OF STUDY

This study directly supports U.S. Army research into the future use of unmanned ground combat vehicles. The results from this study provide insight into whether individual decisions will change with the use of automated or semi-automated systems as part of ground combat teams. This provides a greater context, in general, for human-robotic interactions; providing elements of user trust in automated systems versus trust in live systems. Our results could influence future design concepts as well as future development of doctrine for unmanned systems pertaining to fielding, training, and use.

F. THESIS ORGANIZATION

This thesis is organized as follows: Chapter I presents the problem statement, research questions, scope, and benefits of the thesis. Chapter II provides the background information for unmanned systems in the military, studies in human-robitic integration, studies using virtual environments to evaluate decision making, and the lessons learned from the review of previous literature. Chapter III presents the participant demographics, the experimental design, and the methods for conducting the experiment. Chapter V provides the statistical methods used for analysis, results of the thesis grouped by hypothesis, exploratory information, and post-experiment survey results. Chapter VI discusses the results in summary, outlines the limitations of the study, depicts areas of future work, and wraps everything together with the conclusion. Several appendixes support the work as well as two supplemental documents that will be required for the replication of the study.

THIS PAGE INTENTIONALLY LEFT BLANK

II. BACKGROUND

A. CHAPTER OVERVIEW

The background work for this thesis examines the use of unmanned systems in today's military, studies in human-robotic integration (HRI), tactical level decision making processes, and the use of virtual environments for studies in decision making. We also provide a review of the lessons learned from the literary review that guided how we designed and conducted our experiment.

B. USE OF UNMANNED SYSTEMS IN THE MILITARY

To examine the use of unmanned systems in the military, we will first review the history of unmanned systems and then the growth in trust by the military for unmanned systems.

1. Brief History of Unmanned Systems in the Military

Since 2003, there has been an explosion in the use of unmanned systems in Iraq and Afghanistan (Cruz, 2011). While the term *unmanned system* is used regularly for these combat systems, based on the average interpretation, the systems are currently not completely autonomous (Defense Science Board, OUSD(AT&L), 2012). For this reason, we will refer to them as semi-autonomous, meaning they are controlled by an operator, but have some capacity for completely automated tasks. For many unmanned systems, however, the operator is not generally in the same vicinity as the Soldiers using the system and can often be on a completely different continent. Regardless of the location of the operators, in the author's personal experience gained during three deployments to Iraq, the use of unmanned systems became increasingly prevalent from 2003 to 2010. In Iraq in 2003, unmanned systems were held and used almost solely by brigade levels; and even then, they were limited to large unmanned aerial vehicles (UAVs). In 2006, UAVs had become commonplace at the company level. By the end of 2010, unmanned systems could be found in all aspects of military operations, even at the platoon level. UGVs, ground-based systems, were mostly used by explosive ordinance units to recon or render safe improvised explosive devices (IEDs), but could also be found in some platoon-level units and special operations forces. By 2011, more than 3,000 UGVs and over 400,000 sorties of UAVs had been used by the U.S. Army in Iraq and Afghanistan (Cruz, 2011). By necessity, today's Soldiers have become very familiar with semi-autonomous unmanned systems. Given an increase in these systems, it is very possible that familiarity has garnered trust so that individuals may make the same decisions with live team members as with semi-autonomous systems that have even greater autonomy.

2. Trust in Unmanned Systems

The increase in familiarity with unmanned systems by tactical leaders may have increased the level of trust in autonomous and semi-autonomous systems. Current unmanned systems rely on trust in both the system and the operator, with no real autonomy. From 2003 to 2010, trust was retained mostly with the operator, while trust in the system was limited to trust that the system will function. By 2010, however, there was more evidence that greater autonomy was given to UAVs (Hancock et al., 2011). Increased autonomy, in 2010, allowed UAVs, not UAV operators, to control their flights during simple transit from one safe area to another. The operator would only take control when necessary for more complex actions. The automation afforded to the use of UAVs indicates the link between the level of trust and the degree to which an unmanned system may be employed (Hancock et al., 2011). Given that greater trust may allow for a greater allowance of autonomy, we may expect to see changes in the decisions leaders will make with more advanced autonomous, or semi-autonomous, systems. Leaders may be willing to afford unmanned ground vehicles more autonomy in their actions provided the leaders have a certain degree of trust that the system will operate as expected.

C. STUDIES ON HUMAN-ROBOTIC INTEGRATION

Understanding of the interaction of humans with an autonomous system was gained through an in-depth review of works pertaining directly to human-robotic, or autonomous, integration (Hancock et al., 2011; Lee & See, 2004; Colebank, 2008).

The effects of human, automated systems, and environment-related factors impacting perceived human-robot interaction (HRI) are evaluated and quantified using meta-analytic procedures in a study conducted by the Army Research Laboratory (ARL) (Hancock et al., 2011). In the course of the study, the researchers found several characteristics of the person, the autonomous system, and the environment (see Figure 1) that could affect trust between human and automaton (Hancock et al., 2011).

While personal characteristics and experiences of an individual can affect the degree of trust an individual has in an automated system, we are more concerned with aggregated responses in areas we can control for future system development. Like personal relationships, the degree of trust from one individual to another relies upon previous experience with the individuals and their own psychosocial behaviors (Hancock et al., 2011). By asking participants directly about their degree of trust, we mitigated the effect of the more individual factors that could affect trust within the study. We are focused on the automated and environmental characteristics that shape trust, elements that we can control and that will be relevant to a larger audience.

Consistency, dependability, predictability, and reliability of an automated system will affect a person's trust in the automated system (Hancock et al., 2011). According to the ARL study, "When an individual is not able to predict what an automated or robotic system is supposed to do, trust decreases" (Hancock et al., 2011, p. 6). Trust, by definition, requires that something is reliable (Merriam-Webster, 2014). Therefore, consistency, dependability, and predictability are essential to developing trust in someone or something. A person's trust, however, is given in degrees. The degree of trust that is given is dependent on both reliable performance and societal influence.

The cultural environment in which the automated system is operating can directly affect the degree of trust placed in the system. Studies have found that trust can be defined by the organization and that an individual's trust can be informed by others within the organization (Lee & See, 2004). Also of concern are the methods of communication and interaction between individuals within the organization. Methods of communication and for tasking assignments may also affect trust within the organization (Hancock et al., 2011). Much of this type of trust depends on perceptions of how actions will occur given the communications, tasking, or the like. Trust in someone or something is dependent upon reliable actions performed based on commonly accepted organizational interactions and beliefs. Any variance from what is expected or what is the norm for the organization will affect the degree of trust in whatever is violating those norms.

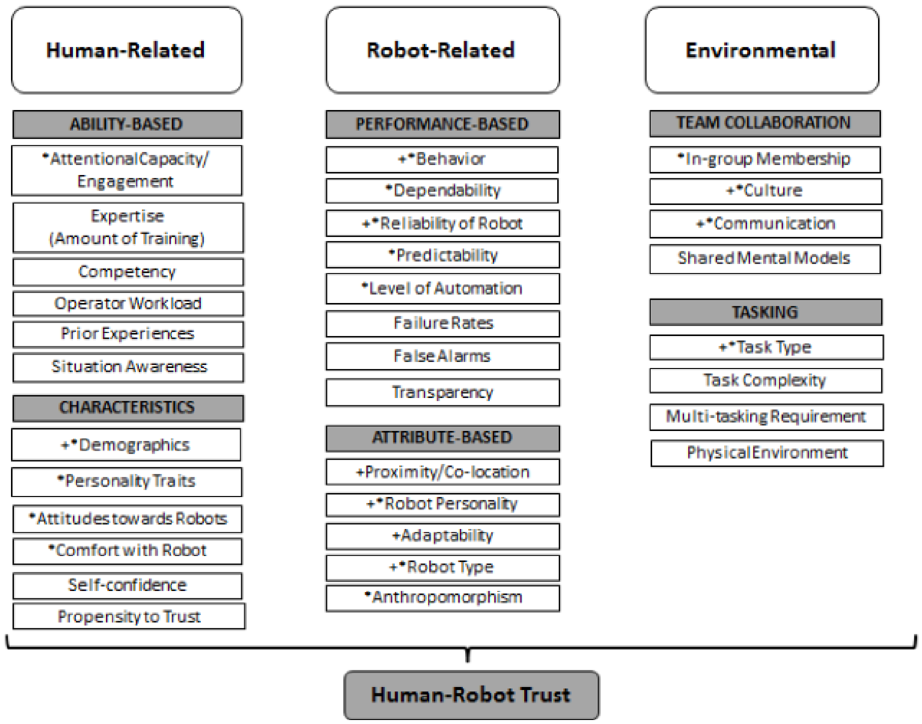


Figure 1. Factors of trust development in HRI (from Hancock et al., 2011)

We may assume that a person's trust in an autonomous system will affect their decisions, especially in a tactical situation where individuals must trust in the actions of their counterparts to keep them alive. Although not a military situation, firefighting has a similar requirement of trust to ensure resolution of a deadly situation. One study explored the differences between human-human teams and human-automaton teams with respect to communications, trust, and efficacy using a computer-based team firefighting game called C3Fire (Colebank, 2008). LCDR Colebank hypothesized that an automated system would change the interactive dynamic, which would result in decreased performance compared to using humans. LCDR Colebank (Colebank, 2008) conducted a between-group experiment in which participants executed a team firefighting simulation. Four participants conducted two simulated scenarios (human-human and human-automaton).

One of the four participants acted as a confederate. The confederate played both scenarios, but was physically present in only one of the two scenarios. In the human-automaton scenario, the confederate played in a separate room, portraying an automated agent within the simulation. Though communication techniques were the same for both the human-human scenario and human-automaton scenario, participants entered the scenarios blind to how the automated agent would actually perform. LCDR Colebank measured trust based on team performance, the density and types of communications, and results of a team efficacy questionnaire. The scenarios were counterbalanced to reduce an order effect.

The results of LCDR Colebank's study (Colebank, 2008) suggest a difference in how automated systems are treated and a decrease in overall performance when using automated systems. While no difference in the volume of messages between human-human teams and human-robot teams existed, the types of messages sent differed. Participants expected more immediate responses and were more likely to override the actions of the confederate who was acting as an automated system. Based on participant comments, LCDR Colebank, concluded that there was far less trust in the automated confederate than the live confederate. Mission performance was decreased in the human-automaton teams compared to the human-human teams. A grouping of success rates (Figure 2) shows that the performances in human-human scenarios were the same regardless of whether they occurred first or second. When the human-automaton scenario occurred first, however, there was a far greater mission failure rate than when the human-automaton scenario occurred second. LCDR Colebank confirms that the differences in performance may have been caused by a participant's lack of knowledge in how the automated agent would operate. By extension, this lack of knowledge may also account for the differences in communication types. LCDR Colebank's results, however, clearly show that differences exist in communication methods, performance, and the degree of trust imparted by participants working in human-automaton teams (Colebank, 2008).

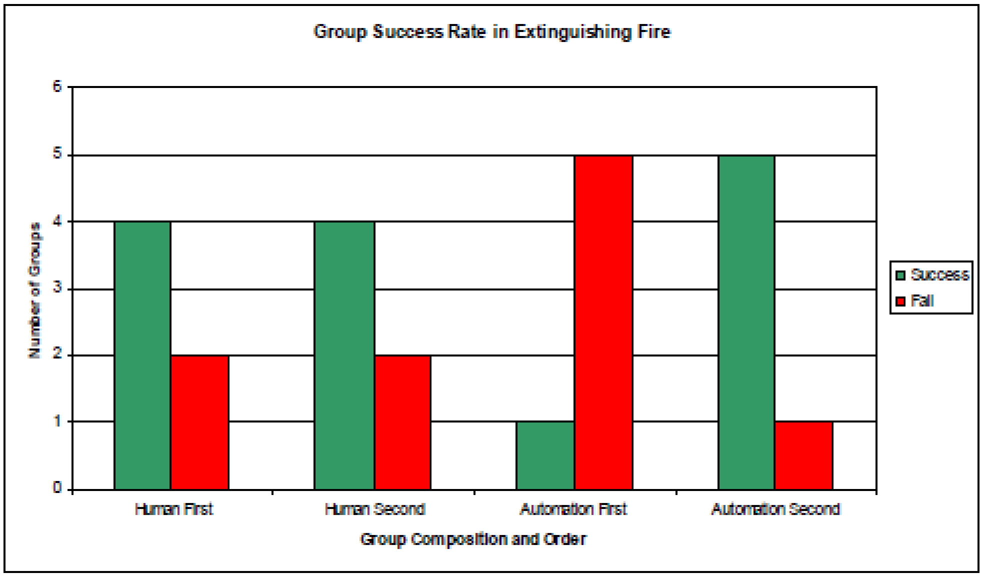


Figure 2. Group success rates in extinguishing fires (from Colebank, 2008)

Studies of human-robotic integration have confirmed that there is a difference in how humans interact with automated systems versus other humans (Colebank, 2008). Human interactions with automated systems rely upon several factors, only some of which are in regard to the automated system itself (Hancock et al., 2011). Environmental conditions such as an organization's workplace culture remind us that an automated system must be able to interact, not just communicate, in a social context (Hancock et al., 2011). Automated systems must inform those it interacts with on what it sees, intends, and can accomplish (Colebank, 2008). An automated system, that is not seen as behaving as expected, will not be trusted. Therefore, an automated system must be introduced to the existing social environment to gain the cultural trust of those with whom it will be working.

D. TACTICAL DECISION MAKING

The tactical level of warfare is "the level of war at which battles and engagements are planned and executed to achieve military objectives, assigned to tactical units or task forces" (Director for Joint Force Development, J-7, 2011, p. 359). Decisions made at the tactical level of warfare are referred to as tactical decisions. Specifically, for the purpose of this thesis, tactical level decisions are those decisions that are made by leaders of tactical units conducting tactical-level operations. Decisions made at this level are generally required to be more reactive and dynamic, with direct feedback either through unit action or opponent reaction. An officer in charge of a four-vehicle platoon consisting of four high-mobility medium wheeled vehicles (HMMWVs) and 20 personnel (five per vehicle) would need to make decisions about how to maneuver the vehicles from a start point to an objective point prior to ever actually moving. We are more concerned, however, with the decisions made during movement, when the platoon leadership must use information from his or her environment to make on-the-spot decisions that may have life-or-death consequences. The immediate life-or-death decisions are what, for the purposes of this study, we consider tactical level decisions.

The Army design methodology used by the U.S. Army to support decision making processes is a cognitive function that uses narrative construction and visual modeling to build a decision-making framework from the individual's experience and knowledge of the situation (ADP 5-0, 2012). Building a decision framework helps individuals understand situations by building mental models to allow a fuller contextual understanding. Decision frameworks are essential to tactical decisions to allow for timely and accurate decisions (ADP 5-0, 2012). Narrative construction and visual modeling are essential for building an effective decision framework.

Narrative construction involves the creation of a story-like background to the situation. It is "the conscious bounding of events and artifacts in time and space" (ADRP 5–0, 2012, p. 2–5). For tactical situations, the narrative construction occurs as part of the first paragraph of the operations order. The first paragraph, or "situation" paragraph, explains the friendly situation, the enemy situation, and the enemy's most probable and most dangerous courses of action. Before a mission, individuals are provided the narrative as part of their mission brief.

Visual modeling is the use of visual graphics to better understand a situation. Like the narrative, this would be done during the orders process as part of the mission brief, prior to the actual mission. The U.S. Army uses terrain models or two-dimensional maps with operational graphics to represent terrain as well as for a visual simulation of the mission during mission rehearsals. During most rehearsals, these visual representations are used to explore all manner of contingencies and mission possibilities to allow more responsive actions from leaders at all levels (ADRP 5-0, 2012). Simulations such as VBS2 provide another dimension of visual modeling. Units have used virtual three-dimensional models and virtual fly-throughs to provide greater visualization of the operation prior to execution.

Leaders use information from the mission brief, terrain model, and rehearsals, along with their own personal experience, to build a decision framework for their actions within a tactical mission. Effectively, they build a list of possible situations and response actions to take in an "if/then" type of methodology. This methodology is supported by the recognition-primed decision (RPD) model (Klein & Calderwood, 1990).

A three-year study by Klein and Calderwood (1990) evaluated the decision-making processes of U.S. Army tank platoon leaders and fire ground commanders in real-life situations. Based on their findings, Klein and Calderwood developed the RPD model (see Figure 3) to explain how decisions are made in tactical-level situations. The first step of the process is to determine if the situation is familiar to the decision maker. In this step, decision makers are comparing the current situations to those they prepared in their decision frameworks. If a situation is not familiar, the leaders will try to gain more information so that the situation matches one prepared in the decision framework. If they cannot find an exact match, then they will attempt to find a "best fit" (Klein & Calderwood, 1990).

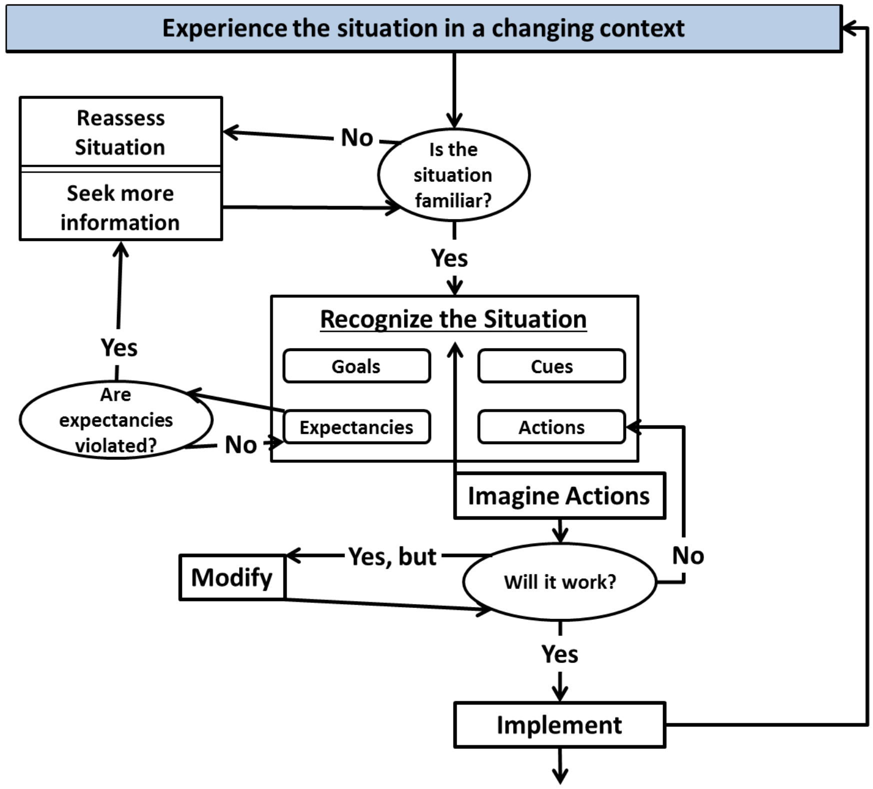


Figure 3. The Recognition-Primed Decision Model (Klein & Calderwood, 1990)

The RPD model was developed from tactical experiences and follows Army doctrine in relation to how decisions are made. The first step is understanding the decision process being used; the next step is developing methods for quantifiably evaluating the decisions made by tactical leaders. The next section describes such a method, situation judgment tests (SJTs). The combination of the RPD method and SJTs can be used to quantitatively evaluate tactical decisions.

E. STUDIES USING VIRTUAL ENVIRONMENTS FOR DECISION MAKING EVALUATION

Situation judgment tests attempt to assess the knowledge, skills, values, or attitudes of individuals by presenting a scenario based on real events and asking the participant to judge and interpret the decisions made within the scenario. SJTs provide a scenario that can be related to actual events or doctrinal methodologies to elicit a participant response that is measured against a standard built from doctrine or expert consensus (Legree & Psotka, 2006).

A technique for creating the measurement standards for an SJT is the critical incident technique. This technique involves the use of experts to report significant events in their areas of expertise and the methods of resolution. By pairing these reports with empirical data and background research, we may be able to identify standards or rules of behavior that were most effective. This technique can also involve using models of human performance and careful deductive reasoning of existing doctrinal actions to evaluate decision success (Legree & Psotka, 2006). The Pacific Science and Engineering Group (PSE) used the critical incident technique in a decision making study with the United States Marine Corps in 2010.

The PSE study used SJTs to evaluate decision making performance both before and after execution of a live training scenario in the Immersive Training Trainer (ITT) located at Camp Pendleton, California (Kobus, Kobus, Ostertag, Kelly, & Palmer, 2010). The ITT blends virtual and live training environments into tactical scenarios for training squad-sized elements. The PSE developed a computer-based SJT that mirrored the live scenarios and then evaluated the computer-based SJT as both a training tool prior to the tactical level scenario and a means for measuring unit performance afterward.

SMEs were used to evaluate tactical level scenarios and minimize their scope to observable behaviors within a live tactical scenario. SMEs validated all possible decision paths or courses of action to ensure scenarios were realistic and robust enough for quantifiable evaluation in both the computer SJT and the ITT. A small group of SMEs were selected from the larger group to take the SJTs. Their scores were then averaged out to determine the expert ratings (correct answers) for the computer-based SJT and the tactical level scenario. The SME scores were used to create a Likert-scale against which the participant's results were compared (Kobus et al., 2010).

The findings from the PSE study show that the use of the computer-generated SJTs prior to the immersive training enhanced small unit performance of complex decision making tasks while providing quantifiable measures of performance for both the immersive environment and the SJT. The techniques used by the PSE group broke down into some simple ideas. Decision making is complex, and there is no one perfectly right answer. The PSE group also validated those choices with SMEs in an attempt to determine an approximate "right choice." This right choice, spread out amongst all of the choices, allowed for a baseline value upon which other choices could be evaluated.

Based on these studies, it appears that virtual environments could be used in cooperation with SJTs to provide a means for testing decision making, provided the scenarios are developed to a rigorous standard based on doctrinal theory and validated by SMEs. It should, therefore, be possible to test if leaders' decisions change when using an unmanned system versus a live counterpart.

1. Using VBS2 to Test Decision Making

Virtual Battlespace Simulation 2 is an Army program of record created by Bohemia Interactive (BI) Group. Bohemia Interactive consists of companies in the U.S., Australia, and Czech Republic. Initially, BI Group was a computer game company, but after the production of Operation Flashpoint, the U.S. Army asked them to create DARWARS Ambush. Over several years, BI Group worked with the U.S. Army, U.S. Marine Corps, and Australian Defense Force to build successively better iterations of Virtual Battlespace Simulation that were more tailored to military simulations and simulators. Recently, Bohemia Interactive was awarded a contract to provide VBS3, an even more developed version. We chose VBS2 because it met all requirements for providing a virtual environment, the project team's previous experience with the simulation and the general within the U.S. Army.

F. LESSONS LEARNED FROM LITERARY REVIEW

Previous studies and literature have provided many key lessons that were helpful for the design of the experiment and data collection techniques. Understanding human-robotic trust, decision making, and methods of data collection were key considerations used for the design and execution of the experiment.

1. Understanding Human-Robotic Trust

ARL's study of human-robotic trust outlined how interactions between human and autonomous systems could increase or decrease trust (Hancock et al., 2011). Primary to the trust relationship was the reliability of the system to do what it was expected to do. LCDR Colebank confirmed this premise, but also outlined an important aspect of executing this kind of experiment—participants must have some expectations for how the autonomous agent will perform (Colebank, 2008). LCDR Colebank's study allowed data to be skewed by people who had no understanding of the autonomous system's capabilities and expected actions other than what would be provided. Therefore, we created a training scenario that trained all participants in the use of the unmanned system. While there was a risk of building too much trust, individual's preconceived ideas and organizational bias would counter the limited familiarity gained from the training scenario (Hancock et al., 2011).

2. Decision Making

Army doctrinal methods for decision making imparted the importance of providing a means for the participant to build a decision framework to be used within the mission scenarios. Key elements for the creation of the framework included having a narrative and visual models (ADP 5-0, 2012). To allow a participant to build an effective decision framework, we provided a situational narrative (see Appendix G) at the start of the scenario along with a two-dimensional map with graphics and unit symbols as seen in Figure 4.

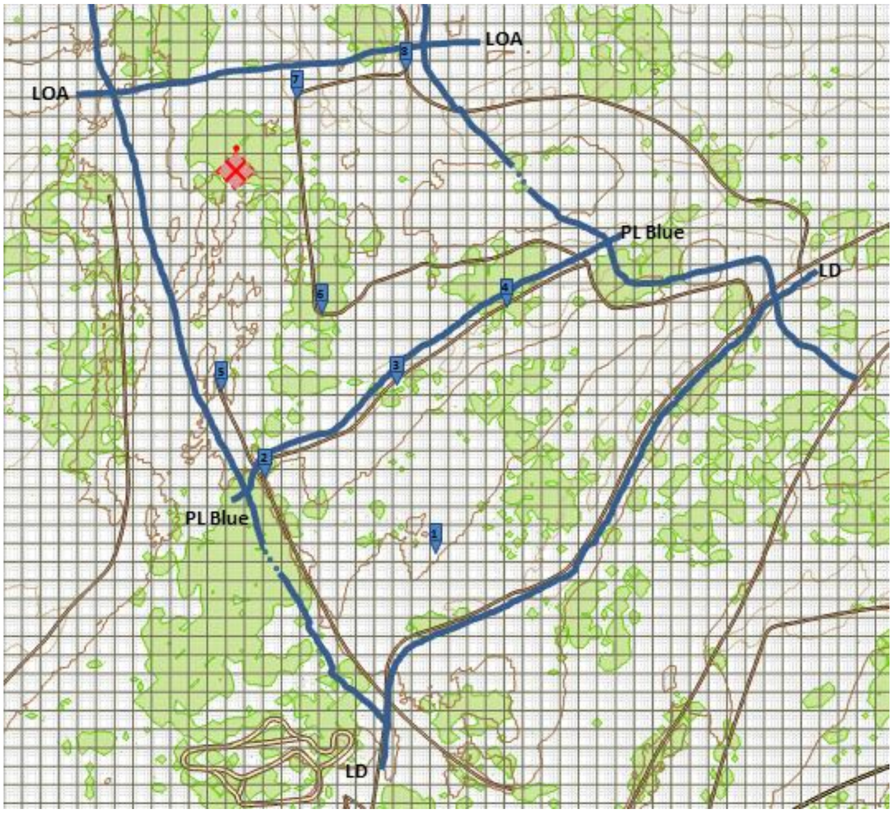


Figure 4. Operational graphic provided at the start of the scenario

The mission narrative, two-dimensional map, and participant's experience allowed for the creation of a decision framework that helped participants execute their decisions.

3. Methods of Data Collection

All the participants' decisions had to be collected and quantified to allow for statistical analysis. The PSE's use of SJTs provided a basis of validity to Legree and Psotka's use of the critical incident techniques for the development of quantifiably evaluated scenarios (Kobus et al., 2010; Legree & Psotka, 2006). The PSE's methods for constraining the scope of its evaluation and Legree and Psotka's use of doctrine and subject matter experts were techniques that wused for this experiment that were invaluable to the creation of scenarios and data collection techniques.

THIS PAGE INTENTIONALLY LEFT BLANK

III. METHODS

A. PARTICIPANTS

Participants of this research were students and faculty of the Naval Postgraduate School with a cross level of U.S. military experience. The project team recruited participants via e-mail, social and organizational network communications (i.e., the NPS muster page), and by word of mouth. Mr. Jesse Huston, a civilian employee, scheduled the participants to avoid any possibility of coercion due to rank or position.

Participants were screened through voluntary completion of a survey form (Appendix B) that assessed their military experience, unmanned systems experience, and individual demographics (age, sex, etc.). Part of the survey included a test (see Appendix C) of the participant's tactical knowledge of Army mounted doctrine. Demographic information was not used to screen out any participants, but was used to block participants and for predictive and qualitative analysis.

Thirty participants completed the experiment from four services: Army (54%), Marine Corps (40%), Navy (3%), and Air Force (3%). Tables 1 and 2 show the participant demographic. No statistically significant differences existed between groups in any of the demographic categories.

DEMOGRAPHICS (Numerical Data)	LIVE (SD)	UGV (SD)	Total Mean
Years of Service	12.03 (3.92)	12.4 (4.87)	12.22 (4.49)
Combat Tours	2.2(1.38)	2 (1.37)	2.1 (1.40)
Length of Tours	8.27 (3.41)	8.9 (3.60)	8.58 (3.58)
Knowledge Test	1.8 (0.98)	2.4 (0.71)	2.1 (0.92)
% of Service Mounted	29.67% (32%)	41.2% (34%)	35% (6%)
% of Unmanned System Experience	17.74% (28%)	10.4% (14%)	14% (22%)
Age	34.07 (4.71)	35 (6.20)	34.53 (5.61)
Experience Rating	216.47 (84.02)	245.37 (55.57)	230.92 (26.64)

Table 1. Demographic characteristics of the participants, mean (SD).

DEMOGRAPHICS (Categorical Data)	LIVE	UGV	Total Count
Sex			
M	14	15	**29**
F	1	0	**1**
Service			
USA	6	10	**16**
USMC	7	5	**12**
USN	1	0	**1**
USAF	1	0	**1**
Specialty			
Combat Arms	8	10	**18**
Special Forces	1	1	**2**
Service Support	2	2	**4**
Combat Service Support	4	2	**6**
Direct Contact Experienced			
Yes	4	5	**9**
Yes - but not leader	2	2	**4**
No	9	8	**17**
Type of Warfare Experienced			
Conventional	0	0	**0**
Low Intensity	6	7	**13**
Both	3	3	**6**
N/A	6	5	**11**
Experience Level			
High	4	4	8
Mid	8	9	17
Low	3	2	5
Mounted Experience			
Both	2	7	9
Tracked	0	0	0
Wheeled	8	4	12
None	5	4	9
Trained in Mounted Combat			
Yes	10	9	19
No	5	6	11
Unmanned System Experience			
Used	0	2	2
Operated	0	0	0
Directed	4	5	9
Used and Operated	1	0	1
Used and Directed	3	2	5
Operated and Directed	0	0	0
Used Operated and Directed	0	2	2
No	7	4	11

Table 2. Demographic information of a categorical nature and displayed by frequency counts

Using the demographic data and the participant's tactical knowledge test score, participants were blocked into experience groups, with random assignment into the live or UGV group, as shown in Figure 5. Experience levels were calculated using a metric of the participant demographic that produced a total numerical value (see Appendix H). Eight participants were placed in the high-experience group (285 and above), 17 in the mid-experience group (157–284), and five in the low-experience group (0–156). Figure 5 shows the breakdown between experience and treatment groups.

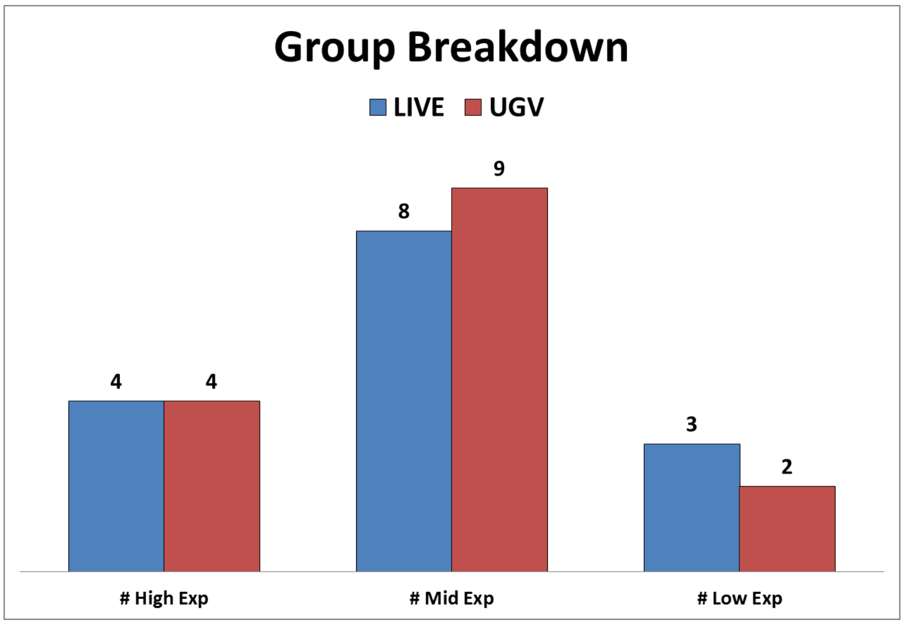


Figure 5. Representation of how participants were blocked by experience and then divided into treatment groups

B. DESIGN

The experimental design for this study relied upon four tenets—control, randomization, replication, and blocking. By using these tenets, we created a design concept (see Figure 6) that drove the methodology of the study, conduct of the experiment, and subsequent analysis techniques. We first blocked participants by experience level and then randomly assigned to either the live wingman or UGV scenarios, and their decisions, decision-making time, and confidence and trust in their wingman were measured.

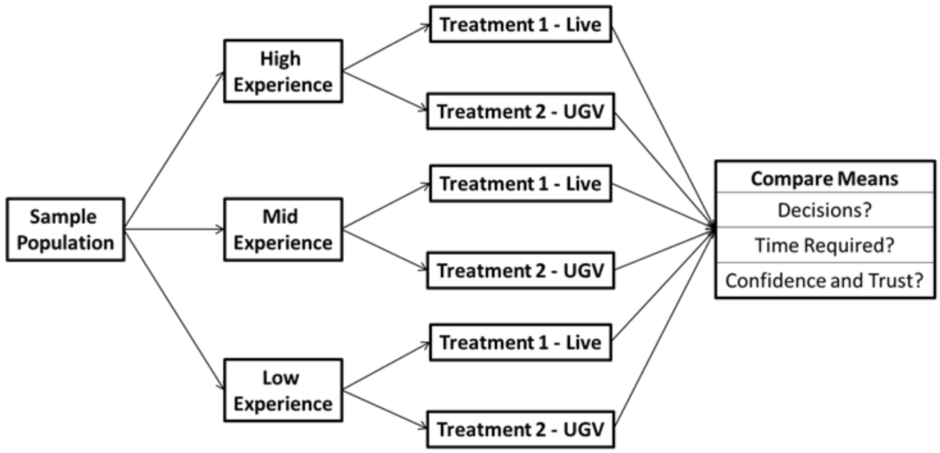


Figure 6. Participant blocking and treatment groups

1. Control

This study involved several sources of variability. Primary sources of variability existed in the participant's decisions, the wingman's actions, and the actions of the scenario's friendly and enemy artificial intelligence (AI).

To control variability from the participants, we developed scenarios that limited their possible decisions to a maximum of four. The goal was to omit possible responses that were outside of the research intent while simplifying the participant's decision process into measurable selections.

We had intended that the participant's wingman for the live scenario should be a human confederate, but in order to control variability in responses between the live and automated wingmen, we opted to use the simulation's artificial intelligence for both the automated wingman and live wingman. By coding in replicated human responses, we were able to provide the appearance of a live counterpart, but with the assurance that the wingman's actions would be the same for every scenario.

We used scripting (a form of computer code for interacting with the simulation) to control the actions of both the enemy forces and the wingman to ensure that behaviors controlled by the simulation's artificial intelligence programming would remain consistent, thus ensuring that all participants received the same cognitive stimulus for their decisions.

2. Randomization

We assigned participants at random to their individual treatments, ensuring that there was no bias in participant groups. We did, however, identify that there may be some variability caused by experience, which necessitated blocking by experience.

3. Replication

The use of a virtual environment allowed for scripting of both the scenario and the wingman's actions to ensure that each participant replicated the same study. The only things that changed between participants were the results of their independent decisions. The scenario and the experience was the same for each participant depending on his or her group. Also, because our study was documented in code, it can be replicated at other locations and facilities to allow for continued experimentation and analysis.

4. Blocking

Using a metric of key demographic data (Appendix H) described above, participants were blocked into experience levels—high, mid, low. The use of a metric ensured that participants could not be assigned due to bias. Participants in each block were then randomly assigned to their treatment groups based on the numbers of individuals needed to ensure equal distribution.

C. SCENARIOS

We designed scenarios within the virtual simulation, VBS2, to approximate a conventional conflict. Scenarios presented participants with a mission to conduct a movement to contact north as part of a Bradley vehicle section, in order to gain contact with enemy infantry forces.

The scenario started with a movement decision. Participants were informed that they would be moving using the travelling technique and were asked who should lead or who should follow. Travelling is the term used for movement when contact is not likely and speed is necessary (Headquarters, Department of the Army, 2001). In the course of their movement, participants encountered an enemy scout vehicle. After killing the enemy scout, the participants had to decide whether to defend from their current position continue movement using the travelling overwatch technique, or continue movement using the bounding overwatch technique. Participants who chose to defend then had to decide which battle positions to occupy and the order in which to move to the positions. Participants who chose to conduct travelling or bounding overwatch then had to decide who should lead the movement. When the movement was complete, the participants were presented with enemy infantry. Once the wingman identified the enemy, the participant had to decide who would provide suppressive fire (the participant or the wingman) and who would assault the infantry position.

1. User Interface

The UGV and live wingman scenarios were coded in such a way that all interactions were limited to those options presented in the HUD and pop-up dialog windows. In general, the HUD allowed participants to communicate with their vehicle gunner or with their wingman in very basic terms. Participants also had the ability to control the commander's independent viewer (CIV) which controlled what they could see on the battlefield. Figure 7 shows a picture of the HUD. Appendix I provides a larger detailed view with descriptions of the components of the HUD.

Participants could command their gunner or wingman to engage targets. The participant's gunner, however, would fire in the direction indicated by the participant's CIV. The wingman would only fire if an enemy threat had been identified. When commanding their vehicle gunner to fire, participants had to click the *fire* command for every three- to five-round burst. When commanding the wingman to fire, the participant only had to give the command once. The participant was able to switch the type of weapon system his or her gunner used from the m240mm coaxial machine gun to the 25mm cannon. The participant could order the wingman to ignore a target if the participant wanted the wingman to stop firing or to scan in other directions. The wingman, however, could never go back to that target again.

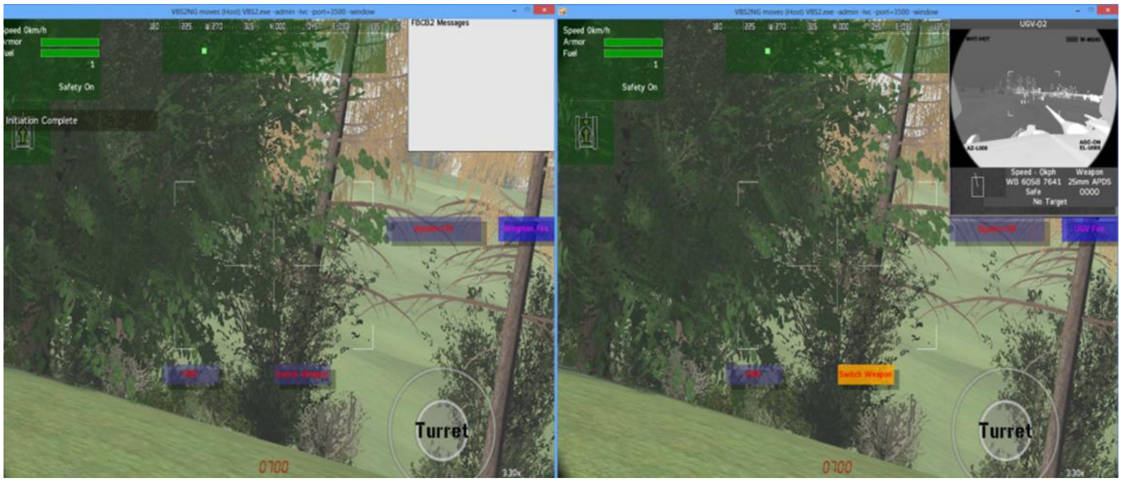


Figure 7. Live and UGV Heads-up-display

In addition to the basic controls, the participants were provided a map with unit icons that updated when vehicles moved within the simulation. The map also displayed an icon for enemy forces, when identified by the wingman. An example of the HUD for the UGV scenario in Figure 8 shows how the map and participant’s view were arranged. Figure 8 also shows the UGV camera that provided the participant a visual of what the UGV saw, some textual information about identified targets and the condition of the UGV’s weapon system.

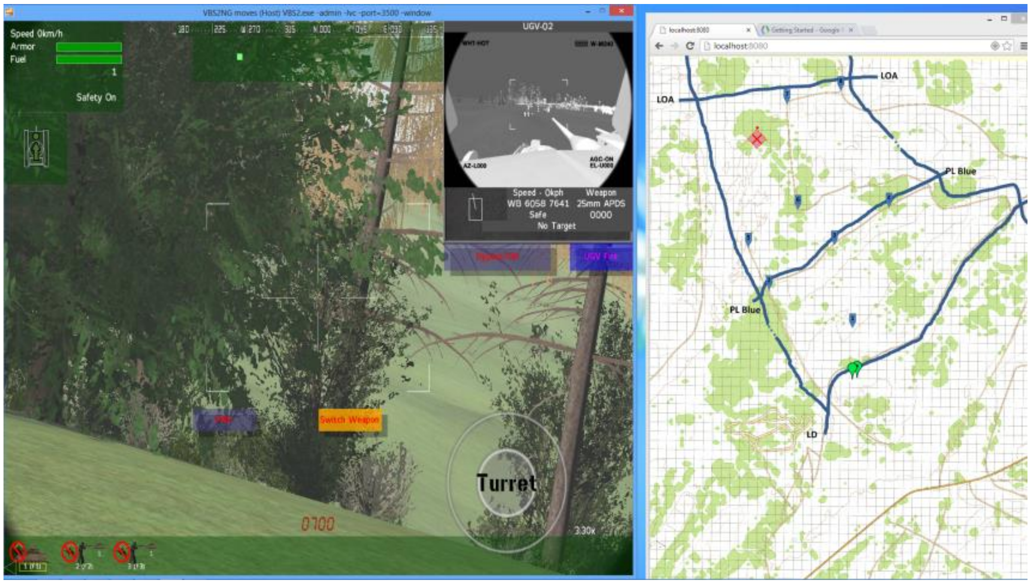


Figure 8. Graphical user interface for UGV Scenario

2. Practice Scenario

The practice scenario provided the user all the general information needed to conduct the scenario. Participants were allowed to conduct the scenario more than once, though only a handful utilized this option. In addition to familiarizing the participant with the basic controls, the practice scenario allowed the user to become familiar with the capabilities of the UGV. Because participants from both the live and UGV scenarios conducted the same practice scenario, both had some relative familiarization with the UGV capabilities from which to answer the post-survey questions.

Participants conducting the practice scenario were first introduced to the interface and provided a guided tour of the various features. They were then allowed to practice engaging stationary targets in order to practice targeting and engaging with their own vehicle gunner. After destroying two targets, one close and one far, participants were introduced to how the UGV would communicate that it had acquired a target (ping noise followed by a zoom into the target by the UGV camera). They were then asked to tell the UGV to engage the target, and they witnessed its ability to accurately engage and destroy an enemy force. Following the engagement, participants were presented a decision dialog screen similar to what would be used in the test scenario. Participants then decided which direction to move and witnessed their section's execution of the movement. For the practice scenario, the UGV always followed.

Once complete, participants were asked if they wanted to conduct the practice scenario again and were assigned their group, either live or UGV. The scenarios for the live and UGV groups were identical except for the methods of communication between the wingman and the participant. Therefore, although the scenarios were identical, some key differences in methods of interaction were scripted into each scenario.

3. UGV Scenario

The UGV's method of communication consisted of a visual representation of the UGV's camera in the upper right corner, textual output displayed under the camera image, and audible pings. The camera provided a constant feed to the participant, informing him or her of where the UGV was looking. In addition, the UGV provided textual input directly beneath the camera. This information presented a GPS location for the UGV and information on any targets, if identified, to include location, target type (if known), and battle damage assessment (BDA). Target information included range, direction, and distance. A small vehicle silhouette provided the participant an idea of the UGV's weapon orientation in relation to the vehicle, and textual data provided information on the current weapon status (armed or safe). Lastly, the UGV communicated to the participant via audible pings. These pings alerted the participant to the fact that the UGV had identified a target.

A fully autonomous system would have been allowed to engage the target on its own. But we decided to require a human to make the final decision to fire in order to be more realistic to the more immediate use of unmanned ground vehicles. Thus, upon identifying an enemy target, the UGV elicited an audible ping and then waited for the user to give the command to fire. The UGV then began engaging the target as directed. In addition to the audible ping, the UGV's camera would also zoom in on the target. Once the target was destroyed, the camera would zoom out to the wider field of view. BDA was then provided in the target status block of the UGV information window.

4. Live Scenario

A single text block and audio radio messages comprised the live scenario interface. In today's modern combat, most message traffic at the tactical level is relayed either with a radio or with the Future Battle Command Brigade and Below (FBCB2). The FBCB2 provides a map image with icons representing various friendly units and tactical messages. FBCB2 allows written communication between individuals on the battlefield. In the live scenario, the textual window in the upper right corner of the participant's HUD provided these textual messages from the faux live wingman.

In addition to the textual messages, radio traffic was simulated by recording the voice of the administrator. The voice recording was then scheduled to play at key moments in the scenario, such as when enemy threats were identified or had been destroyed. Participants, however, were informed that while they would receive the wingman's radio messages, their own communication would be limited to the HUD and the clickable buttons on the screen. The intent of the audio radio traffic was to provide a more convincing "live" scenario. A few participants attempted to shout back at the wingman, indicating some success in the ruse.

Although the methods of communication were key components of both the UGV and live scenarios, they were only a means to immerse the participant into the scenario—live or UGV. In addition to representing the live or UGV scenarios, we had to incorporate a means for requesting, capturing, and measuring the participants' decisions.

D. MEASURES

We utilized quantifiable methods to measure a participant's response within the scenario as well as their survey questions in the post-experiment survey. Each participant's decisions and opinions could then be analyzed against the total sample using statistical methods to compare results.

1. Scenario Decision Measures

We rated decisions to provide an individual path score for each participant. The sum of all the decisions generated the path score.

a. Path Score

Each scenario involved four decisions broken down into two basic categories—movement or tactical. A movement decision involved the participant's choosing to either lead or follow during a movement, regardless of what type of movement (travelling, travelling overwatch, or bounding overwatch). A tactical decision required a decision about how to employ the participant's section to achieve a goal. Depending on the decisions the participant made, he or she would move down a specific path of actions. Figure 9 shows that the paths are fairly straightforward, except for the ability to select more information. Participants could request more information, which afforded them more time to make a decision. By selecting more information, participants were given 10 additional seconds to scan the battlefield with their CIVs and 40 additional seconds to make a decision. This option, however, could only be selected four times in total. Very few participants utilized this feature, and those who did tended to think they would receive another pop-up with even more information.

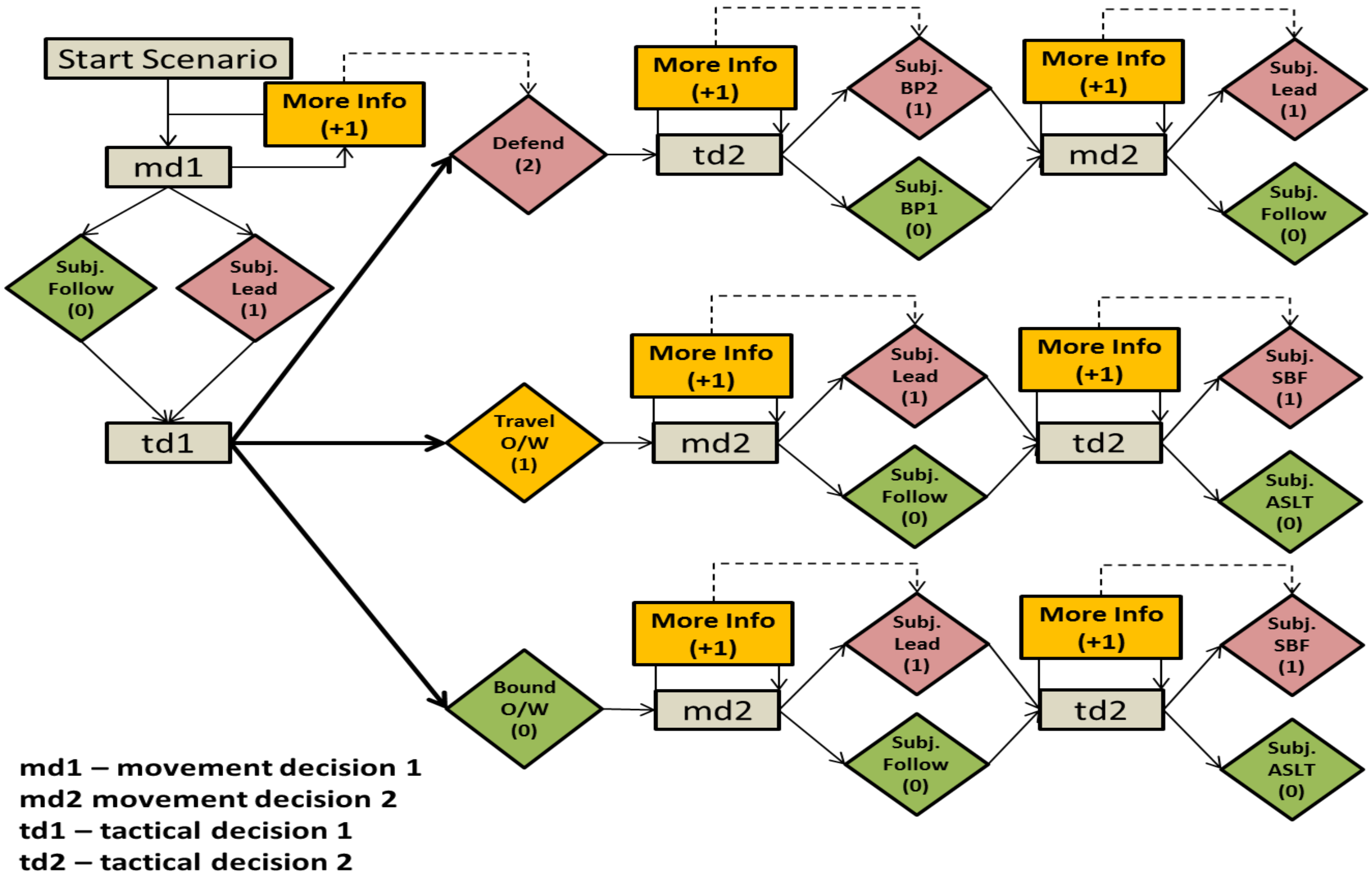


Figure 9. Decision path and scoring

In order to conduct a quantifiable comparison of decisions, a value, or score, had to be attributed to each decision. Decisions were scored based on a review of doctrine and consensus from SMEs. If a participant chose the expected decision, his or her score for that decision would be zero. The more the participant diverged from what was expected, the higher the score would be for that decision. The sum of participant's scores for each decision is referred to as the path score. The path score was the overall measure of decision performance.

b. Movement Decision One

Movement decision one was the first decision for the participant, requiring the participant to decide who would lead and who would follow when using the movement technique of "travelling." Participants received a 0 if they chose to follow and a 1 if they chose to lead. At the end of the movement, participants encountered an enemy scout that they had to destroy. The participants' instructions to their gunners to fire and any command to their wingmen to fire were recorded using a time stamp and an indicator of the action taken.

c. Tactical Decision One

Participants chose to either defend or move into one of two movement techniques (travelling overwatch or bounding overwatch) following their encounter with the enemy scout. Travelling overwatch is used when contact is possible but speed is necessary. Both vehicles move, but one moves in such a way as to ensure that it can provide relative security and early warning for the other. Bounding overwatch is used when contact is expected; it consists of the two vehicles conducting either alternating or successive bounds. One vehicle provides security and early warning while the other vehicle moves. Once the vehicle that was moving is set in a good position, it provides security so that the vehicle that had been providing security can move. Participants' who chose to defend received a 2, travelling overwatch received a 1, and bounding overwatch received a 0. Following their selections, participants either moved to the defensive positions or continued their movement forward using the movement technique selected.

d. Movement Decision Two

During movement decision two, those who choose defend decide who will provide security first, while the other vehicle moves to its defensive position (battle position). Participants who chose to move to their position first received a 1, while those who chose to let their wingmen occupy their position first received a 0.

Participants who chose to conduct travelling overwatch or bounding overwatch had to choose both who would lead and who would follow for the next movement. Scoring was the same for this decision as movement decision one, so that those who chose to follow would receive a 0 and those who chose to lead would receive a 1.

e. Tactical Decision Two

Those who chose to defend must then choose who would occupy each of the two battle positions. The google map provided participants information about where the positions were located and gave them an idea of the terrain. Battle position 1 was more restrictive than the other and was more at risk of infantry attack. Battle position 2 had longer fields of view and less risk. Those who chose battle position 1 for themselves received a 0, and those who chose battle position 2 received a 1.

After the first bound, those who chose to conduct travelling or bounding overwatch made contact with enemy infantry moving in the tree line to their front. Participants decided if they wanted to act as the support-by-fire or the assault element. The support-by-fire element suppresses or destroys the enemy force with long-range aimed fires while the assault element maneuvers to a position that allows it to close with and destroy the enemy force. Those who chose the support-by-fire position received a 1, while those who chose the assault position received a 0.

2. Time to Make Decision

Computer system time tracked when the scenarios began and ended. In addition, to calculate the time between decisions, each decision was time stamped with the computer system time. By calculating the time between the start and finish, we were able to generate the time the participant took to complete the scenario. In addition, we could calculate the time used to make a decision by calculating the time difference between the each decision.

E. EQUIPMENT

The equipment used for this work included software and hardware that allowed for both the conduct of the experiment and the analysis. Software used to execute the experiment included the Virtual Battlespace Simulation 2 program, Netbeans, Websocket Gateway, and Global Mapper. Hardware used to conduct both the experiment and the analysis included an Alienware computer, monitor, and the necessary input devices. Software tools for data collection and post analysis of results included notepad, JMP Pro 10, and Microsoft Excel.

1. Software

Virtual Battlespace Simulation 2, Netbeans, Websocket Gateway, and Global Mapper were used to create the total scenario experience. VBS2 provided the virtual environment. Netbeans provided the basis from which to code the websocket gateway and the means to run the websocket gateway code. The websocket gateway acted as a means to translate data from VBS2 to the google map page. Global Mapper helped to convert from the VBS2 map to modified google tiles.

a. Virtual Battlespace Simulation 2

VBS2 is part of the Army's Games for Training program of record (McCaney, 2014). It is a 3D first-person networkable simulation that provides realistic semi-immersive environments over large and dynamic terrain areas with hundreds of simulated military and civilian entities (Milgaming, 2010). The terrain consists of both geo-specific and geo-typical terrains. Models within VBS2 consist of the latest Army, Marine Corps, and Navy equipment. VBS2, which comes with an integrated scenario editor, allows for scripting actions and scenarios and can communicate via distributed interactive simulation (DIS) and high-level architecture (HLA) protocols (Milgaming, 2010). VBS2 is an Army-adopted training medium that is familiar to the author, which made it a good fit for the project. The networkable characteristics and ability to script scenarios provided a diverse platform from which to run a controlled experiment. Milgaming provided training in how to administer VBS2 and some instruction in basic scripting techniques was provided by Milgaming. In addition, Milgaming also provided the code for the UGV HUD which was then modified for the live HUD by the project team. All other support for utilizing and scripting in VBS2 came from the VBS2 wiki provided by Bohemia Interactive Simulations (Bohemia Interactive Simulations, 2014).

b. Netbeans and Websocket Gateway

Websocket Gateway is an open source program created by Professor Don McGregor at the Naval Postgraduate School. The code was slightly modified to allow interaction with VBS2 DIS protocol traffic and Google Maps. The Websocket Gateway utilized websockets to create a gateway between VBS2 and the Google Maps interface. Netbeans was utilized to run the Websocket Gateway and adjust the code for the use within the experiment.

c. Global Mapper

Global Mapper is included with the VBS2 developer suite and allows for interaction with 2D maps with 3D data. Exporting the VBS2 map for the scenario in a TIF format, Global Mapper was used to convert the map into Google tiles for representation in the Google interface and as part of the Websocket Gateway.

2. Hardware

An Alienware computer, model Aurora R4, consisting of an Intel® Core™ i7–3930K CPU with 32 GB RAM running at 3.20GHz utilized a 64-bit Microsoft Windows 8 Operating System to facilitate the experiment. VBS2 and Netbeans were loaded onto the computer and all additional programs were built into the system. A standard keyboard and mouse provided a means for user input, though participants only utilized the mouse. The display was provided by a 32-inch LED HDTV (720p, 60Hz) manufactured by Sceptre.

3. Data Collection Systems and Software

Notepad, Microsoft Excel 2010, and JMP Pro-10 were utilized to store and conduct statistical analysis on data. All files exported from VBS2 were in text format (.txt) and were read by Notepad. Files were stored by subject ID in folders similarly labeled. All demographic data was transferred immediately from demographic surveys and the experience test prior to the commencement of the scenario. Formulas within Excel allowed immediate calculation of experience levels. Excel maintained a running count of all participants assigned to each experience level by group. Following the completion of the scenario, all data from the output text file was transferred to Excel either manually or by importing text data. Initial data counts and preliminary analysis was computed in Excel and then verified in JMP. JMP Pro 10, 64-bit edition, was utilized for the majority of statistical calculations and graphical representations. JMP Pro is the advanced analytics version of JMP, providing visual data access and manipulation, as well as guided support for the methods and results of statistical calculations.

F. SURVEYS

All participants conducted survey's to support the experiment. Demographics survey's conducted prior to the experiment provided basic information about the participants and aided in determining how the participant would be blocked. A tactical knowledge test conducted prior to the experiment allowed a means to measure a participant's knowledge of Army doctrinal terms that were used in the experiment. This further supported the blocking of the participant by experience. Post surveys conducted after the experiment provided participant insights into their own decisions and their feelings about their wingman.

1. Demographic Survey

The demographic survey included typical demographic questions along with specific questions that could indicate a participant's experience level in both mounted combat and unmanned systems (Appendix B). Some of the key questions included

a. Have you had any experience with unmanned vehicles?

b. What kind of mounted vehicle experience do you have?

2. Tactical Experience Test

The tactical knowledge test assessed an individual’s knowledge of the Army-defined doctrinal terms utilized within the scenario (Appendix C). Three multiple choice questions comprised the test:

a. What is a movement to contact?

b. Of the three movement techniques, Travelling, Travelling-Overwatch, Bounding-Overwatch; what are the main differences and when are they used?

c. Using several pictures of various formation types, indicate at which position the leader should be.

3. Post-Experiment Survey

The post-experiment survey captured why participants made their specific decisions, their overall confidence in their decisions, and their level of trust in their wingman following the scenario. Participants were asked to indicate their level of trust in the wingman’s accuracy of information, expected actions, and knowledge of situation. Lastly, participants were asked if they would trust an automated system in the future and to explain why (Appendix D).

G. PROCEDURES

In order to conduct our study and ensure reliable results, we had to implement strict procedures to avoid errors by issues.

1. Control Procedures

To eliminate extraneous variables, control procedures used in this study included scripted directions, standardized mission briefings, scripted computer-based simulations, blocking by experience, and random assignment to treatment groups.

2. Before the Experiment

Prior to the arrival of the participant, the participant's name was listed next to a confidential list of subject IDs and participant contact information. After the information was recorded the document was secured. From that point on, the participant was only referred to by his or her subject ID. After determining the subject ID to use, administrators initiated the primary scenario computer and all necessary programs. A second computer was turned on outside of the lab area and VBS2 was started on this computer as well. The participant walked past the second computer when he or she entered the room. The computer's presence and visible use of VBS2 was meant to further impart the belief that the wingman for the live scenarios was a real person. Once all computers and software were running, individual subject folders were created in both hard copy and digital formats. A hard copy folder was utilized for all surveys and the consent form. A digital folder was established on the computer and appropriate naming conventions (subject ID) were verified in the scenario script allowing export to the digital folder. Subject IDs were updated in the Microsoft Excel data file and checked against existing subject IDs to verify that no duplicates occurred.

Upon arrival, participants were greeted and thanked for their participation. Following introductions, they were briefed from an initial script which outlined the experiment, its purpose, and the volunteer nature of the study. All participants were then allowed to read through the consent form (Appendix F) before signing. After signing the consent form, participants filled out the demographic survey and tactical knowledge test. The participants experience was determined using their demographic survey and tactical knowledge test. Participants were blocked by their experience into a groups(high, mid, or low experience) and then randomly assigned to a treatment (UGV or live) within their assigned group.

Participants were provided a practice scenario so they could become familiar with both the user interface and the unmanned ground vehicle before conducting the mission scenario. All participants went through the same practice scenario and utilized the UGV interface. After the participants became familiar with the interface and practice scenario, the administrator reloaded the scenario to the main menu and reinstituted the Google Map interface and Websocket Gateway to ensure these programs were operational for the current scenario.

A copy of the mission brief for the scenario was laminated to the table in front of the participant. The participant was informed to read the mission brief prior to conducting the scenario. In a very few cases, the mission brief was given verbally as well. In all cases, the administrators answered any questions about the mission. If conducting the live scenario, the participant was informed that the wingman would be one of the administrators and that the administrator would operate the wingman vehicle from a computer in the adjoining room. When both the administrator and the participant were ready, the participant was instructed to select his or her specific scenario type as indicated in Figure 10.

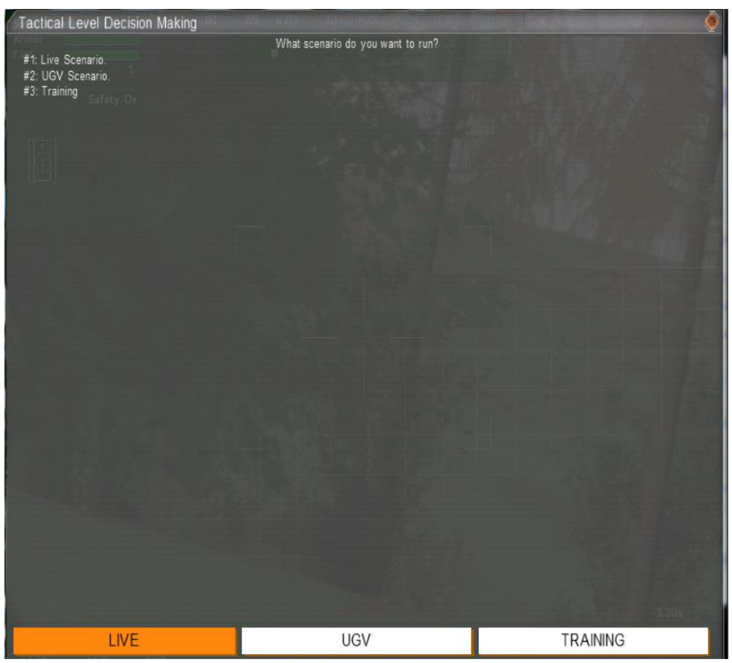


Figure 10. Main menu provided scenario selections of Live, UGV, or Training.

3. During the Experiment

For those conducting the UGV scenario, one or both administrators were always present. For those conducting the live scenario, one administrator remained in the room while the other stayed outside in an attempt to further build the belief that the wingman was a live person. The participant navigated through the scenario utilizing a mouse at each decision point within the scenario. After starting their particular scenarios (live or UGV), the participants were provided information that contact was not likely for the first 5 kilometers during their movement to check point 1 (CP1), after which contact would become likely. The participants were then asked if they would like to lead or follow (see Figure 11).

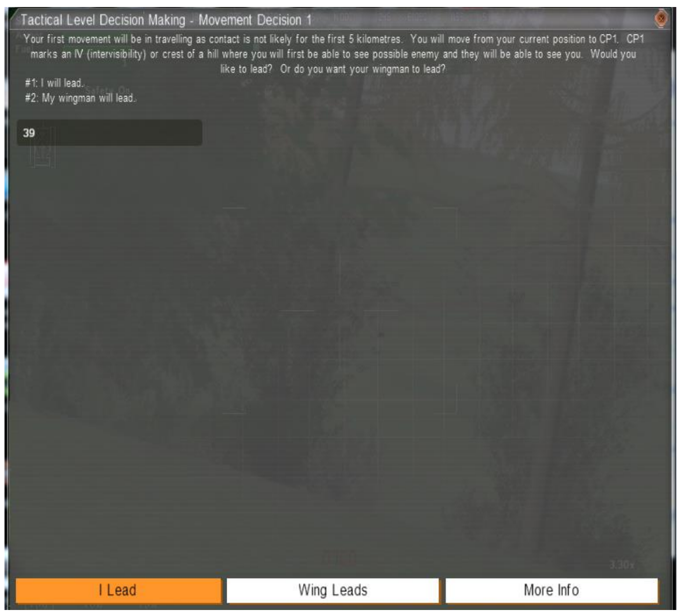


Figure 11. Movement Decision One: lead or follow wingman

Participants then scanned around their positions as the vehicles maneuvered based on the participants' decisions. The lead vehicle slowed its movement to allow for the trail vehicle to catch up prior to reaching CP1; this allowed both vehicles to come online prior to cresting the hill. Upon cresting the hill, an enemy scout vehicle was identified. Depending on the participant's previous decision, the scout may have been identified first by the wingman or participant. Once the scout was identified, the participant either took action to engage the enemy scout vehicle from his or her vehicle, had the wingman engage, or both (see Figure 12).

Figure 12. Enemy Scout identified by participant and UGV in UGV scenario

After the enemy scout was destroyed, the pop-up window for tactical decision one was presented to the participants. The participants were informed that they had made contact with enemy reconnaissance and asked if they wanted to defend, transition to the travelling overwatch movement technique, or transition to the bounding overwatch movement technique (see Figure 13).

Figure 13. Tactical Decision One: defend, travelling overwatch, bounding overwatch

If participants chose to defend, they were provided another pop-up window, tactical decision two, that asked if they wanted to defend from battle position 1 (BP1) or battle position 2 (BP2). BP1 and BP2 locations were related using check point graphics available on the participants' maps (see Figure14).

Figure 14. Tactical Decision Two, defense path: BP1 or BP2

Once the participants selected one of the two options, they were then provided the movement decision two pop-up. Participants were asked if they wanted to move first while the wingman provided overwatch, or if they wanted the wingman to move first while they provided the wingman overwatch (see Figure 15). After this selection, the scenario was ended.

Figure 15. Movement Decision Two, defense path: lead or follow

If participants chose either travelling overwatch or bounding overwatch, they were then provided movement decision two. Movement decision two asked the participant to decide who should lead the movement or who should follow (see Figure 16). After making their selections, the participants again scanned while their vehicles moved according to their decision.

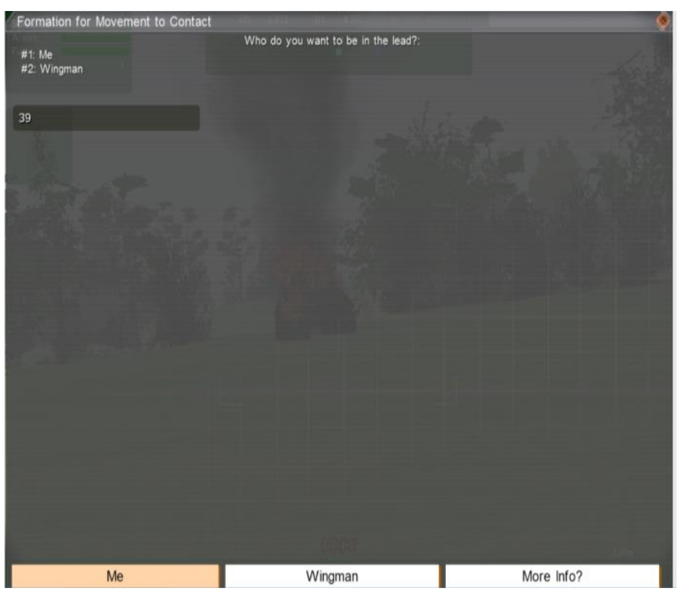


Figure 16. Movement Decision Two, non-defense path: lead or follow

Once both vehicles were online on the other side of the hill, just past the dead enemy scout, the participant and his or her wingman were able to identify enemy infantry. In the majority of cases, the wingman identified enemy infantry first regardless of movement technique. Once the wingman identified the enemy, the final decision was presented: tactical decision two. Tactical decision two asked the participants if they would like to assault the enemy dismounts themselves while their wingmen provided support-by-fire or vice versa (see Figure 17). Following their selections, the scenario was ended.

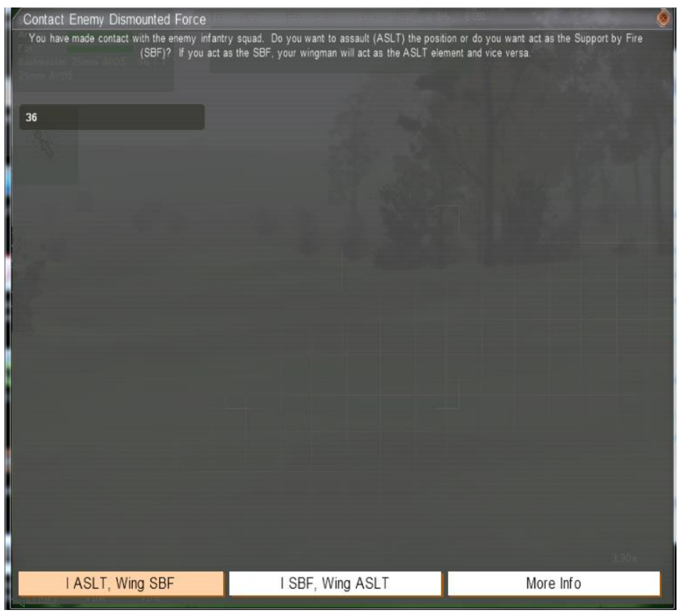


Figure 17. Tactical Decision Two for non-defend paths: Assault or support by fire

4. After the Experiment

After the last decision, the participant was provided his or her final score. Regardless of how close or divergent the participants were in their final path scores, they were informed that they did a good job. Prior to the post-survey, administrators filled in the types of decisions made at each decision point and then verified that what was observed was what was stored in the digital subject folder. Participants were then provided the post-survey and given as much time as needed to complete it. An administrator would then look over the information and ensure completion before conducting a debriefing. All participants were debriefed on the use of deception in the study, whether they were deceived (live group) or not (UGV group). Debriefing also consisted of reviewing the importance of the study and how the participant's involvement supported the research. A basic graphical depiction of results by group was provided for those participants who were interested in the data collected,. Participants were asked not to inform anyone else as to the deception and were provided a copy of their consent forms prior to departure. All participants were thanked for their participation.

IV. RESULTS

A. STATISTICAL METHODS

Prior to conducting analyses, we checked the assumptions and conditions for each statistical method. Both the UGV and live group were independent as was the data for each group. The total sample size was sufficiently large (15 per group, 30 total). Individual characteristics of data for each hypothesis, however, required different statistical methods and evaluation for each.

Path score was the sum of all the decisions, resulting in an ordinal data type. We used the Wilcoxon Rank Sum to test the ordinal values between the two groups (live and UGV). Individual decision data and post-test survey data were categorical, lending to use of Chi-Square tests. Some factors within the decision data had too low a count to provide any relevant feedback (see Table 3). For those decisions, the outliers were excluded from the data.

	MD1		TD1			MD2		TD2			
	LEAD	FOLLOW	DEFEND	TRAV OW	BOUND OW	LEAD	FOLLOW	SBF	ASLT	BP1	BP2
LIVE	7	8	0	8	7	4	11	9	6	0	0
UGV	9	6	1	4	10	7	8	5	4	1	0
	MD1 - Movement Decision 1						**TRAV OW** - Travelling Overwatch				
	TD1 - Tactical Decision 1						**BOUND OW** - Bounding Overwatch				
	MD2 - Movement Decision 2						**SBF** - Support by Fire				
	TD2 - Tactical Decision 2						**ASLT** - Assault				

Table 3. Frequency of decision selections by type.

Decision times, confidence in decisions, and trust in wingman comprised continuous values, and therefore the two sample *t*-test was used to test between group differences in decision time. Where sample standard deviations between groups were not sufficiently close, Welch's Test was used to account for the variance. Whenever there were outliers, we conducted statistical analysis with and without outliers to determine if there was statistical significance with or without the outliers.

Preliminary Results

Figure 7 provides a representation of decision paths by frequency and flow with the percentage of participants who chose various paths. The most common path was to lead the initial movement, then transition to bounding overwatch after initial contact, either follow or lead before assuming the support-by-fire position to assist the UGV's assault.

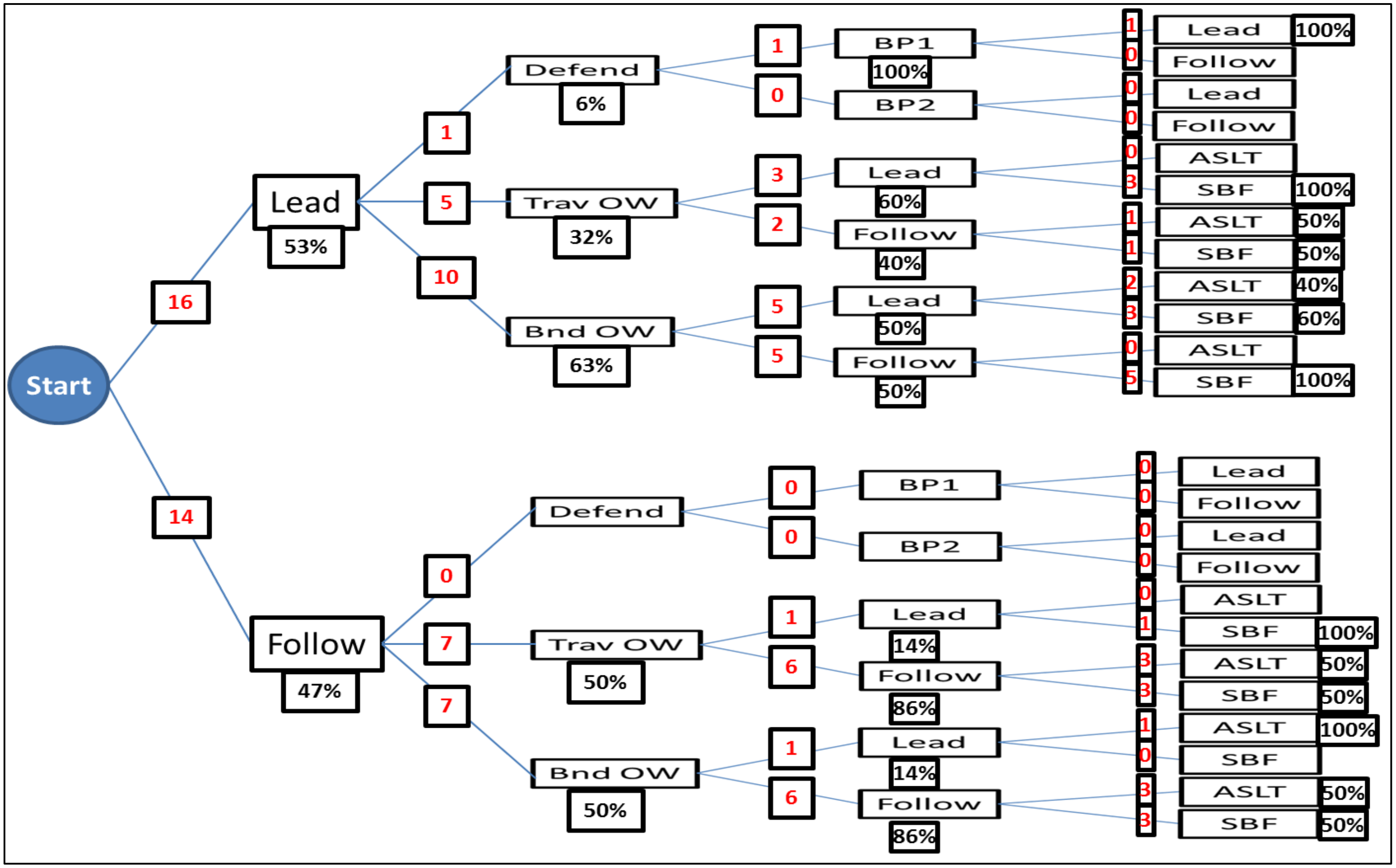


Figure 18. Decision path selections

B. HYPOTHESIS 1—THE MEAN ACCURACY OF EACH GROUP'S DECISIONS WILL BE THE SAME.

We first tested Hypothesis 1 by examining whether group differences occurred for the global measure of decisions, the path score. Next, group comparisons for each of the five decisions were made.

1. Path Scores

Figure 19 shows that there was no statistically significant difference in path scores between the live and UGV groups (z = 0.34, p = 0.73).

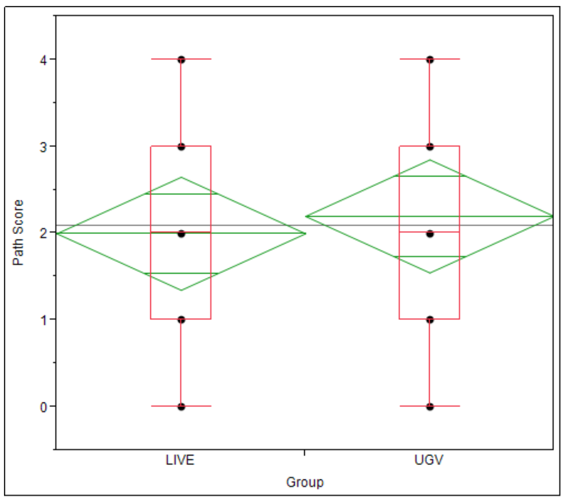


Figure 19. Path score by group

2. Movement Decision One

There was no statistical significance between groups ($X^2(1)$ = 0.537, p = 0.464). Additionally, there was no statistical significance between those who chose to lead (n=16) and those who chose to follow (n=14), regardless of group (Z = 0.232, p = 0.591).

	Follow		Lead			
LIVE	count	8	count	7	ROW Count	15
	total %	26.67%	total %	23.33%	ROW Total %	50.00%
	col %	57.14%	col %	43.75%		
	row %	53.33%	row %	46.67%		
UGV	count	6	count	9	ROW Count	15
	total %	20.00%	total %	30.00%	ROW Total %	50.00%
	col %	42.86%	col %	56.25%		
	row %	40.00%	row %	60.00%		
	COL Count	14	COL Count	16		
	COL Total %	46.67%	COL Total%	53.33%		

Table 4. Contingency table for movement Decision One, comparison between groups

3. Tactical Decision One

One outlier, a single participant who chose to defend, was excluded from the analyses (see Figure 21). There was no statistically significant difference in tactical decision one between travelling overwatch and bounding overwatch ($X^2(1) = 1.857$, $p = 0.1730$). There was no statistically significant difference in the proportion of people who chose bounding overwatch ($n = 17$) versus travelling overwatch ($n = 12$), regardless of group ($Z = 0.66$, $p = 0.745$).

	Bounding Overwatch		Travelling Overwatch			
LIVE	count	7	count	8	ROW Count	15
	total %	24.14%	total %	27.59%	ROW Total %	51.72%
	col %	41.18%	col %	66.67%		
	row %	46.67%	row %	53.33%		
UGV	count	10	count	4	ROW Count	14
	total %	34.48%	total %	13.79%	ROW Total %	48.28%
	col %	58.82%	col %	33.33%		
	row %	71.43%	row %	28.57%		
	COL Count	17	COL Count	12		
	COL Total %	58.62%	COL Total%	41.38%		

Table 5. Contingency table for tactical Decision One, comparison between groups

4. Movement Decision Two

There was no statistical significance ($X^2(1) = 1.304$, $p = 0.253$) between groups for their decisions in movement decision two (Figure 22). The trend across all participants was toward following ($n = 19$) vice leading ($n = 11$) ($Z = 1.03$, $p = 0.063$).

	Follow		Lead			
LIVE	count	11	count	4	ROW Count	15
	total %	36.67%	total %	13.33%	ROW Total %	50.00%
	col %	57.89%	col %	36.36%		
	row %	73.33%	row %	26.67%		
UGV	count	8	count	7	ROW Count	15
	total %	26.67%	total %	23.33%	ROW Total %	50.00%
	col %	42.11%	col %	63.64%		
	row %	53.33%	row %	46.67%		
	COL Count	19	COL Count	11		
	COL Total %	63.33%	COL Total%	36.67%		

Table 6. Contingency table for movement Decision Two, comparison between groups

5. Tactical Decision Two

There was no significant difference ($X^2(1) = 3.84$, $p = 0.70$) between groups for their decisions in tactical decision two (see Figure 23). There was a trend ($Z = 1.03$, $p = 0.063$) by all participants to choose support-by-fire ($n = 19$) rather than assault ($n = 11$).

	Assault		Support by Fire			
LIVE	count	6	count	9	ROW Count	15
	total %	20.00%	total %	30.00%	ROW Total %	50.00%
	col %	54.55%	col %	47.37%		
	row %	40.00%	row %	60.00%		
UGV	count	5	count	10	ROW Count	15
	total %	16.67%	total %	33.33%	ROW Total %	50.00%
	col %	45.45%	col %	52.63%		
	row %	33.33%	row %	66.67%		
	COL Count	11	COL Count	19		
	COL Total %	36.67%	COL Total%	63.33%		

Table 7. Contingency table for tactical Decision Two, comparison between groups

6. Exploratory Analysis Regarding Hypothesis 1

Although the sample size was too small to statistically analyze results by experience, examining the data with descriptive statistics by experience levels can provide worthwhile feedback that may be used in future studies.

a. Experience and Path Score

When reviewing total path score, those with higher experience levels scored generally better (lower overall path scores) than those of lower experience (higher overall path scores) (see Figure 24).

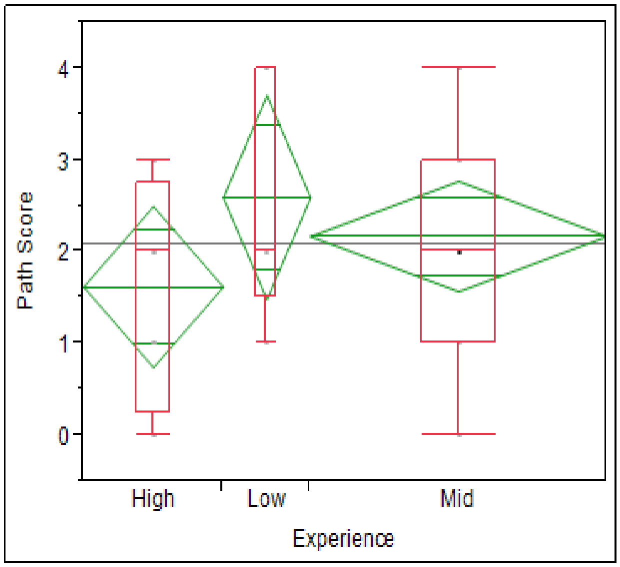


Figure 20. Average Path Score by experience level

When path score is examined by experience levels by group, high experience level path scores remained relatively unchanged, regardless of whether they were in the live or UGV group; mid-level tended to be higher for the UGV group; and low level did not really change; but had less standard deviation (see Figure 25).

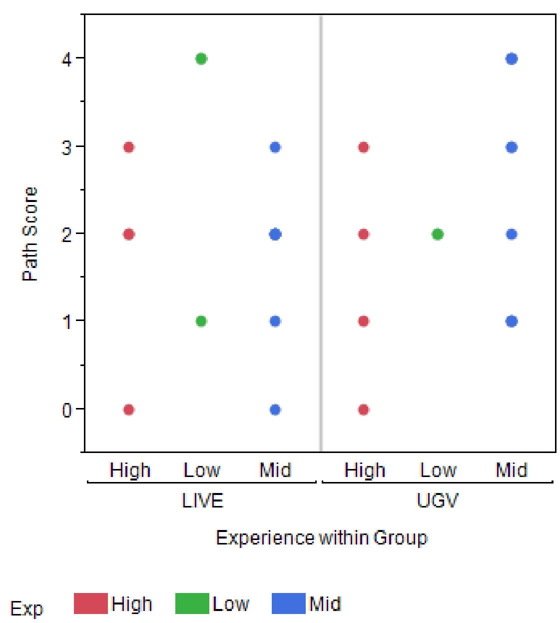


Figure 21. Path score by experience and group

b. Decisions and Experience Levels

Comparisons of specific decisions by experience levels demonstrate interesting differences that may be used for future studies (see Table 4).

In movement decision One, higher-experienced participants preferred to follow their wingman during the initial movement, whereas low and mid-experienced participants chose to lead the movement. Although 10 of 16 mid-level experienced participants preferred bounding overwatch in tactical decision one, the high- and low-level experienced participants were split approximately evenly between bounding overwatch (0) and travelling overwatch (1). High- and mid-level experienced participants preferred to follow their wingmen in movement decision two, whereas low-level experienced participants chose to lead. In tactical decision two, low- and mid-level experienced participants preferred support-by-fire, whereas high experienced participants choose to conduct the assault.

Movement Decision 1

Experience	Count Total % Col % Row %	0	1	
	High	5 17.24 35.71 62.50	3 10.34 20.00 37.50	8 27.59
	Low	2 6.90 14.29 40.00	3 10.34 20.00 60.00	5 17.24
	Mid	7 24.14 50.00 43.75	9 31.03 60.00 56.25	16 55.17
		14 48.28	15 51.72	29

Tactical Deciion 1

Experience	Count Total % Col % Row %	0	1	
	High	4 13.79 23.53 50.00	4 13.79 33.33 50.00	8 27.59
	Low	3 10.34 17.65 60.00	2 6.90 16.67 40.00	5 17.24
	Mid	10 34.48 58.82 62.50	6 20.69 50.00 37.50	16 55.17
		17 58.62	12 41.38	29

Movement Decision 2

Experience	Count Total % Col % Row %	0	1	
	High	6 20.69 31.58 75.00	2 6.90 20.00 25.00	8 27.59
	Low	2 6.90 10.53 40.00	3 10.34 30.00 60.00	5 17.24
	Mid	11 37.93 57.89 68.75	5 17.24 50.00 31.25	16 55.17
		19 65.52	10 34.48	29

Tactical Decision 2

Experience	Count Total % Col % Row %	0	1	
	High	4 13.79 40.00 50.00	4 13.79 21.05 50.00	8 27.59
	Low	1 3.45 10.00 20.00	4 13.79 21.05 80.00	5 17.24
	Mid	5 17.24 50.00 31.25	11 37.93 57.89 68.75	16 55.17
		10 34.48	19 65.52	29

Table 8. Contingency tables for movement decisions by experience level

C. HYPOTHESIS 2—THE MEAN TIME FOR A DECISION BY EACH GROUP WILL BE THE SAME

There were three outliers for time to complete. The beginning time and end times for one participant (Outlier A) could not be recorded, so times are lacking for movement decision One, tactical decision two, and scenario completion time. Two other participants had long decision times on movement decision One (Outlier B) and tactical decision two (Outlier C) that were well beyond three standard deviations from the mean times for those decisions. However, as Outliers B and C did not occur because of technical issues, their data is considered to be valid. Therefore, times to complete the scenario and individual decision times were calculated with and without the respective outliers.

1. Overall Time to Complete Scenario

There was no significant difference in the time to complete the scenario between groups with Outlier A excluded ($F(1,13.6) = 1.177$, $p = 0.297$), with Outlier A and B excluded ($F(1,15.0) = 0.1549$, $p = 0.699$), or with Outlier A, B, and C omitted ($F(1,24.6) = 1.101$, $p = 0.304$).

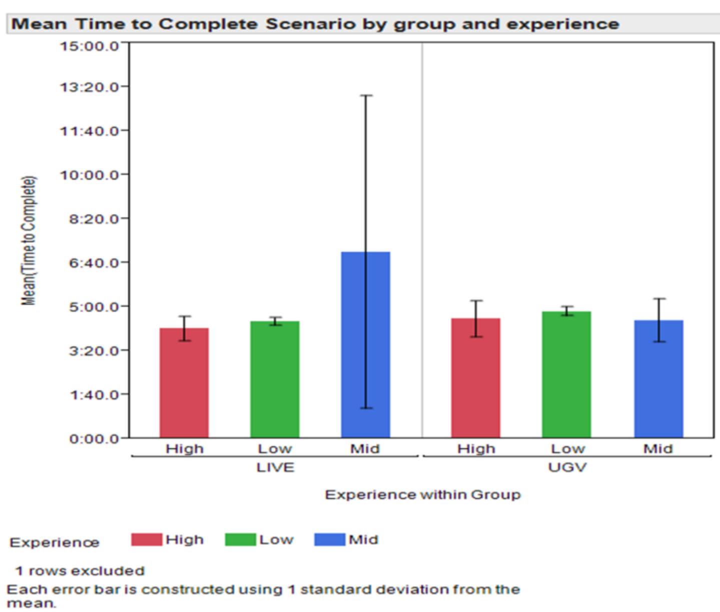


Figure 22. Mean time to complete scenario by group and experience

2. Time for Each Decision by Group

There was no statistically significant difference between any decision time and experience, regardless of whether or not outliers were included (all p’s > .30) (Table 9).

Decision time by decision and group					
	Movement Decision 1	**Tactical Decision 1**	**Movement Decision 2**	**Tactical Decision 2**	**Time To Complete**
LIVE	1.348 (4.152)	2.427 (0.195)	0.093 (0.053)	1.798 (0.437)	5.888 (4.584)
UGV	0.350 (0.142)	2.447 (0.342)	0.159 (0.191)	1.589 (0.139)	4.544 (0.706)
	1 Outlier Omitted			**1 Outlier Omitted**	**Time to Complete (MD1 Omitted)**
LIVE	1.444 (4.292)			1.926 (1.680)	4.759 (1.852)
UGV	0.350 (0.142)			1.589 (0.537)	4.544 (0.706)
	2 Outliers Omitted			**2 Outliers Omitted**	**Time to Complete (TD2 Omitted)**
LIVE	0.297 (0.118)			1.496 (0.503)	5.527 (4.559)
UGV	0.350 (0.142)			1.589 (0.537)	4.544 (0.706)
					Time to Complete (2 Outliers Omitted)
					4.274 (0.632)
					4.544 (0.706)

Table 9. Mean time and standard deviation (SD), in minutes, by group and decision

3. Exploratory Analysis Regarding Hypothesis 2

Table 10 depicts the mean decision times by level of experience. When the outliers are removed, decision times did not vary by experience level (all *p*’s > .30).

Decision time by decision and experience					
	Movement Decision 1	**Tactical Decision 1**	**Movement Decision 2**	**Tactical Decision 2**	**Time To Complete**
High	0.300 (0.159)	2.452 (0.188)	2.452 (0.188)	1.315 (0.707)	4.381 (0.583)
Mid	1.256 (3.892)	2.441 (0.337)	2.441 (0.337)	1.847 (1.563)	5.704 (4.184)
Low	0.343 (0.134)	2.403 (0.156)	2.403 (0.156)	1.777 (0.298)	4.593 (0.249)
	1 Outlier Omitted			**1 Outlier Omitted**	**Time to Complete (MD1 Omitted)**
High	0.343 (0.111)			1.502 (0.505)	4.381 (0.583)
Mid	1.256 (3.892)			1.847 (1.563)	4.775 (1.740)
Low	0.343 (0.134)			1.777 (0.298)	4.593 (0.249)
	2 Outliers Omitted			**2 Outliers Omitted**	**Time to Complete (TD2 Omitted)**
High	0.343 (0.111)			1.502 (0.505)	4.381 (0.583)
Mid	0.313 (0.145)			1.493 (0.573)	5.399 (4.122)
Low	0.343 (0.134)			1.777 (0.298)	4.593 (0.249)
					Time to Complete (2 Outliers Omitted)
					4.381 (0.583)
					4.388 (0.820)
					4.593 (0.249)

Table 10. Mean time and standard deviation (SD), in minutes, by experience and decision

D. HYPOTHESIS 3—THE CONFIDENCE LEVEL OF EACH GROUP WILL BE THE SAME.

1. Confidence by Group

There was no statistically significant difference ($t(28) = 0.533$, $p = 0.598$) in participant confidence in their decisions between groups (see Figure 32). Regarding experience level and confidence, the average confidence seemed to coincide with a participant's experience level. Individuals of higher experience tended to be more confident.

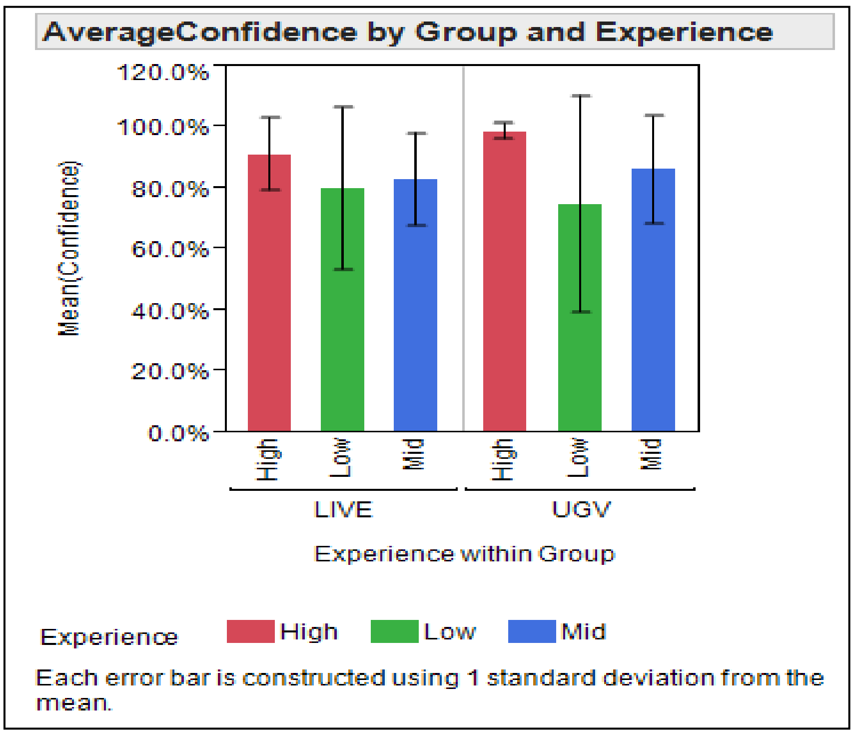


Figure 23. Graphical representation of confidence by group for each level of experience

E. POST-EXPERIMENT SURVEY RESULTS

Three questions from the post survey were analyzed to gain insights into participants' level of trust in their wingmen and what aspects (expected actions, knowledge of situation, and accuracy of information) of their wingman they trusted most or least.

1. Percentage of Trust in Wingman by Group

There was no significant difference ($t(28) = 0.900$, $p = 0.376$) in the percentage of trust in the wingmen between groups. Although there was no significant difference, the average trust for the UGV group was lower than that of the live group.

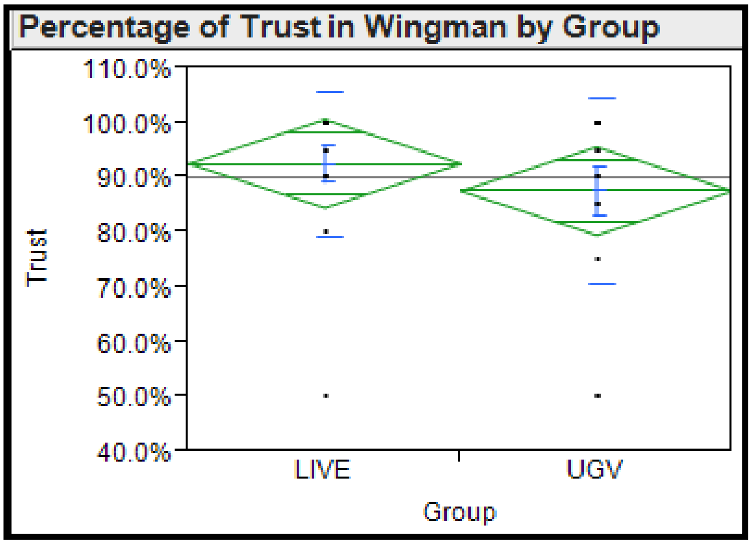


Figure 24. Average percentage of trust in wingman by group

2. Trust in Autonomous Systems

There was no significant difference ($X^2(1) = 0.687$, $p = 0.407$) between groups in whether or not a participant trusted an autonomous system (see Figure 34). Regardless of group, the trend was to trust the wingman ($Z = 1.81$, $p = 0.0702$).

Trust Autonomous System?

Group	Count Total % Col % Row %	No	Yes	
	LIVE	3 10.00 37.50 20.00	12 40.00 54.55 80.00	15 50.00
	UGV	5 16.67 62.50 33.33	10 33.33 45.45 66.67	15 50.00
		8 26.67	22 73.33	30

Figure 25. Trust in autonomous systems (Yes or No) by group

3. Aspects Most or Least Trusted in Wingman by Group

There was a statistically significant difference ($X^2(2) = 11.826$, $p = 0.003$) in the aspects of a wingman that were trusted most between groups. The live group was relatively evenly distributed between all three aspects (accuracy of information, expected actions, knowledge of situation), whereas a majority of the UGV group's responses indicated that they trusted that the UGV would act as expected (see Figure 35).

Aspect Trusted Most about Wingman

Group	Count Total % Col % Row %	Accuracy of Info	Expected Actions	Knowledge of Situation	
	LIVE	5 16.67 55.56 33.33	4 13.33 26.67 26.67	6 20.00 100.00 40.00	15 50.00
	UGV	4 13.33 44.44 26.67	11 36.67 73.33 73.33	0 0.00 0.00 0.00	15 50.00
		9 30.00	15 50.00	6 20.00	30

Figure 26. Aspects trusted most about wingman by group

There was a statistically significant difference ($X^2(2) = 7.650$, $p = 0.022$) in the aspects of a wingman that were trusted least between groups. The live group was relatively evenly distributed between all three aspects trusted least, though the group appeared to lean toward expected actions and knowledge of situation as being least trusted. The majority of the UGV group's responses indicated that they did not trust the UGV's knowledge of the situation (see Figure 35).

Aspect Trusted Least About Wingman

Group	Count Total % Col % Row %	Accuracy of Info	Expected Actions	Knowledge of Situation	
	LIVE	3 10.00 75.00 20.00	6 20.00 85.71 40.00	6 20.00 31.58 40.00	15 50.00
	UGV	1 3.33 25.00 6.67	1 3.33 14.29 6.67	13 43.33 68.42 86.67	15 50.00
		4 13.33	7 23.33	19 63.33	30

Figure 27. Aspects trusted least about wingman by group

4. Exploratory Analysis Regarding Post-Experiment Survey Results

As Figure 28 depicts, there was no significant difference between experience levels for the percentage a participant's wingman was trusted. However, participants of mid-level and low experience leaned toward trusting autonomous systems whereas high experience participants were evenly distributed (see Figure 29). When we examine the data regarding aspects most or least trusted in wingmen, we see that high- and low-level experienced participants most frequently chose expected actions as what they trusted most and knowledge of situation as what they trusted least. Mid-level participants most frequently chose accuracy of information as what they trusted most and selected knowledge of situation as what they trusted least in their wingmen (see Figure 30).

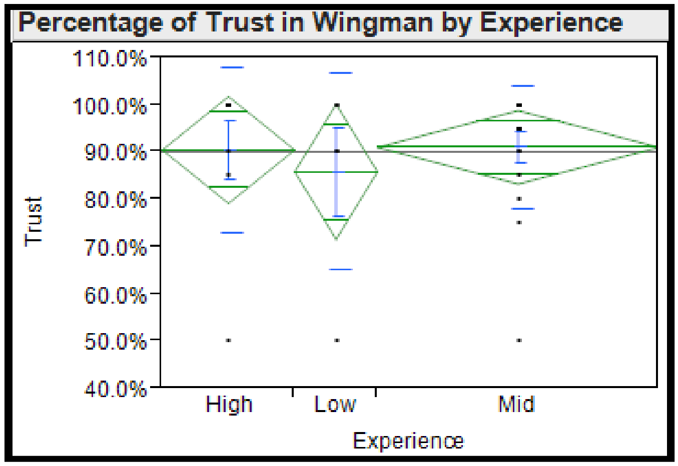


Figure 28. Average percentage of trust in wingman by experience level

Trust Autonomous System?

Experience	Count Total % Col % Row %	No	Yes	
	High	4 13.33 50.00 50.00	4 13.33 18.18 50.00	8 26.67
	Low	0 0.00 0.00 0.00	5 16.67 22.73 100.00	5 16.67
	Mid	4 13.33 50.00 23.53	13 43.33 59.09 76.47	17 56.67
		8 26.67	22 73.33	30

Figure 29. Trust in autonomous systems (Yes or No) by experience

Aspect Trusted Most about Wingman

Experience	Count / Total % / Col % / Row %	Accuracy of Info	Expected Actions	Knowledge of Situation	
	High	1 3.33 11.11 12.50	5 16.67 33.33 62.50	2 6.67 33.33 25.00	8 26.67
	Low	0 0.00 0.00 0.00	4 13.33 26.67 80.00	1 3.33 16.67 20.00	5 16.67
	Mid	8 26.67 88.89 47.06	6 20.00 40.00 35.29	3 10.00 50.00 17.65	17 56.67
		9 30.00	15 50.00	6 20.00	30

Aspect Trusted Least About Wingman

Experience	Count / Total % / Col % / Row %	Accuracy of Info	Expected Actions	Knowledge of Situation	
	High	1 3.33 25.00 12.50	1 3.33 14.29 12.50	6 20.00 31.58 75.00	8 26.67
	Low	1 3.33 25.00 20.00	1 3.33 14.29 20.00	3 10.00 15.79 60.00	5 16.67
	Mid	2 6.67 50.00 11.76	5 16.67 71.43 29.41	10 33.33 52.63 58.82	17 56.67
		4 13.33	7 23.33	19 63.33	30

Figure 30. Aspects trusted most and least about the participant's wingman by experience

THIS PAGE INTENTIONALLY LEFT BLANK

V. DISCUSSIONS

A. OVERVIEW OF STUDY AND SUMMARY OF RESULTS

The purpose of this study was to determine if tactical leaders make different types of decisions when using an automated wingman versus a live wingman. The study attempted to use a virtual environment to present a participant with a conventional small unit or tactical combat scenario in which the participant was required to make critical decisions about how to maneuver and fight using either a live or an automated wingman. The UGV used in the scenario was semi-autonomous; a term used in the study to mean that key critical tasks required a human-in-the-loop for execution. The live wingman was a computer represented entity that reacted in the same way as the UGV to the decision stimulus of the participant. Participants were also blocked by level of tactical experience.

Results regarding decisions made, time to make decisions, and level of confidence in decisions consistently indicated no significant differences between live wingmen and UGV wingman groups. The lack of statistical significance suggests that a greater level of trust may exist toward automated systems than previously thought: Participants with a UGV wingman made the same decisions, primarily for the same reasons, in the same amount of time with the same level of confidence as participants with a live wingman.

Indeed, survey results indicate that individuals are willing to trust autonomous systems. This willingness, however, is extremely contingent upon meeting several criteria. Of the 22 participants who answered that they would trust an autonomous system in the context that it was presented; only one indicated that he would trust an autonomous system with no caveat. Fifteen of the 22 who said they would trust an autonomous system said they would only trust an unmanned ground vehicle in limited missions that did not require complex decisions or actions, and they would require human-in-the-loop vice purely autonomous systems. Although leaders' decisions did not differ when using an automated system in this scenario, further tests are required to see if decisions differ when an automated system is used in a more complex situation, such as an urban environment.

1. Implications of Exploratory Analysis of Experience

In general, the higher a participant's experience level, the better his or her path score. This did not, however, translate to all high-experienced participants returning the expected responses for every decision. Typically, higher experienced individuals did tend to choose to follow in their movement decisions, but they were split on whether to conduct the expected decision of bounding overwatch or the less-expected decision of travelling overwatch. Higher experienced individuals also tended to choose to conduct the assault. The divergence from expected responses, then, was mostly due to the split in bounding or travelling overwatch, at least for higher experienced individuals. Those who had mid-level experience tended to choose to lead initially, but then would choose to follow after contact with the enemy. They also tended toward bounding overwatch and would more often choose to act as the support-by-fire element in an assault on enemy infantry. Low-experienced participants tended to choose to be the lead throughout the scenario and were split between bounding and travelling overwatch. Low-experienced individuals also choose to act as the support-by-fire.

Reasons for decisions were relatively the same between experience groups, usually referring to an attempt to maintain command and control in accordance with the participants' understanding of doctrine or as a means for ensuring situational awareness for them or their wingmen. The largest area of difference was in bounding and travelling overwatch. How a person interpreted the mission greatly impacted the decision to conduct bounding or travelling overwatch. Most participants who chose travelling overwatch indicated that the reason for their choice was due to an attempt to maintain speed, thus necessitating travelling overwatch versus bounding overwatch. Regardless of their choices, most cited doctrine as their reason for the choice they made, which necessitated a change in formation. The difference between bounding and travelling overwatch came down to an interpretation of whether speed was needed or not. The one discrepancy in tactical decision two was a single participant who chose to conduct the defense. The participant chose the defense versus travelling or bounding overwatch because of his interpretation of the mission. Because he thought that contact with a mounted platform meant he had a company-sized element to his front, he chose the defense as a means to allow his superiors to provide a new assessment and send reinforcements. Again, this participant's decision was based solely on his interpretation of the mission and doctrinal response rather than any bias toward an unmanned ground vehicle. In fact, there was only one respondent who provided distrust in the automated wingman as a reason for his decisions. This individual, who was in the high-experience group, chose to lead in both movement decisions, chose bounding overwatch, and chose to conduct the support-by-fire. Because the sample size was so small, we cannot know if this was a trend across more experienced participants that may have a significant effect on how automated ground systems are utilized in the future.

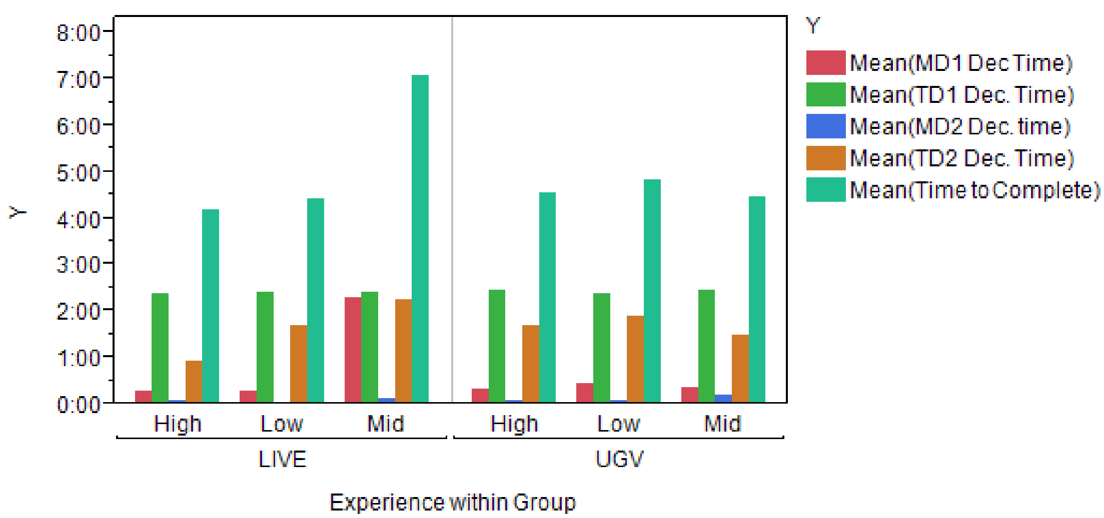


Figure 31. Mean decision times by group and experience

2. Should VBS2 be Used for Testing Future Tactical Concepts Regarding UGVs?

Using VBS2 to conduct an experiment like this one is relatively new. Therefore, it can be of benefit to explore what were found to be the benefits and drawbacks of using VBS2 in an academic study.

The benefit of VBS2 was that it was relatively easy to use for this application. The scripting characteristics were difficult for a novice programmer, but the online resources provided by both Bohemia Interactive and Milgaming made it relatively easy to learn and execute the code. Having an automated scenario ensured replication of events every time for every participant while also providing a means to electronically capture actions and decisions.

The main drawback of VBS2 is that in order to allow for a more dynamic scenario, heavy coding would be needed that most likely would go beyond the capabilities of the general programmer. In addition, because a scripting language is being used, glitches sometimes develop as additional code is piled on top of the existing VBS2 code. Sometimes the code created for a mission may not interact well with the existing game engine or updated versions, requiring rewrites and causing system faults. Because licensing is restricted to the computer for which it is requested, it is difficult if the setup has to be changed partway through execution.

B. LIMITATIONS

The study had several limitations due to the time required to conduct the study and the research team's lack of experience with preparing the computer-based simulation. Some of the limitations included surveying confidence prior to the start, the sample size, the use of a faux live wingman, and the lack of a complex mission scenario.

1. Participant Confidence

Participant confidence was not measured prior to the conduct of the study. Because the degree of confidence a person has in their self can indicate their degree of confidence in an autonomous system (Hancock et al., 2011), it should be included in a any future studies and was a limitation to the conduct of this study. This information could have been used to better weigh the final confidence rating and as a means to determine the effects of individual confidence on trust in automated systems for military purposes.

2. Sample Size

Although no significant differences often existed between groups, interesting trends by experience and group*experience emerged. However, due to the lack of sample size within experience levels, these results could not be statistically evaluated. A larger sample size should be used to allow for a determination of differences due to experience. Furthermore, the sample consisted of a narrow margin of military leaders attending NPS. Using a larger military population than the one at NPS to create a more diverse sample may also allow blocking by military specialty to assess the types of decisions that might be made by a logistician versus a cavalry officer or by a non-commissioned officer compared to an officer.

3. Live Wingman

Although the faux live wingman provided predictability and repeatability of behaviors, it also may allow participants to have doubts about whether the wingman was truly live. The study administrators took great pains to mislead the participants into thinking the wingman was live; however, the lack of actual radio conversation between the two may have mitigated some of the realism in the live group responses.

4. More Complex Scenario

As has been stated, there was trust in the wingman primarily because the situation was very simple and straightforward. There were no civilians on the battlefield and no friendly forces forward of the participant or his or her wingman. Any contact the automated wingman made would be with an enemy force. Therefore, to get a better understanding of individual decisions and the degree of trust people have in autonomous systems, a more complex scenario should be developed requiring participants to make decisions that tests their trust in the UGV's knowledge of the situation, the component they trusted the least in this study.

In addition to increasing the complexity in the scenario itself, the user interface should be improved to allow more freedom to maneuver, meaning that participants can control the movement of their vehicles and that of their wingmen rather than have the vehicles travel on a set path in between decisions. As conducted, the experiment was more like a theme park ride, with participants making very limited decisions and then riding the tracks until they came to another decision point. Giving the participant the ability to control their movement and that of the UGV would allow greater complexity as well as offer a measure of trust by the participant in the automaton's path planning capability. This scenario could be accomplished by using the 2D map interface to select routes or by networking a driver, gunner, and passenger into the scenario. The driver would maneuver the vehicle as per the commander's direction, the gunner would engage targets as directed, and the passenger would control the movements and activities of the UGV in a human-in-the-loop interface.

5. Truly Autonomous Systems

The use of a virtual environment allows an opportunity to test truly autonomous capabilities early before development. This study should be expanded to provide a complete autonomous capability that will allow future autonomous system developers to flush out what interfaces, information, and tools are required for leaders to use and work with unmanned, completely autonomous systems.

C. AREAS OF FUTURE WORK

The purpose of this study was to determine how leader's decisions differ depending on if they have a live wingman or an automated wingman. This study was conducted based on a ground combat scenario utilizing a desktop simulation. There are many opportunities for others to carry this work forward.

Decision making is not just a tactical task. There may be differences in how leaders of larger units assign mission tasks due to the use of unmanned vehicles. The assignment of unmanned systems can overlap into many different high-level command structures for various military services. For example, a similar VBS2 simulation could be used by the Marine Corps to explore how a Marine Expeditionary Force commander might use unmanned amphibious vehicles to conduct an amphibious assault. The Navy could explore how decisions are made by a ship commander when he or she has the use of unmanned submersibles. The use of unmanned systems within the military is growing and there is a need to understand how others might use those systems and how the use of those systems may affect their decisions in the future.

Further studies should be conducted in decision making and what information is required to make those decisions. By utilizing eye tracking and an electroencephalogram, insight could be gained into what type of information is actually stimulating a person's decisions. The use of both those capabilities within this study could have provided information about what participants looked at in the UGV window to determine their decision choices. Likewise, these capabilities could be used in any number of simulations that provide a visual stimulus to determine trends in what people look at to make a decision and what their brain activity suggests about their personal state (sleepy, engaged, etc.) when they make the decision.

Finally, this study could be expanded to explore group decisions with UGVs in which one crew or team member is controlling the UGV. By networking a similar scenario, an actual vehicle crew could be replicated complete with a UGV operator. The crew would then have input on how the UGV was used, and the leader would not be giving commands to an automated system, but to a human who is controlling the automated system. The networked crew could then be coupled with another networked crew so that two crews and two UGVs could replicate an augmented platoon formation of four tanks, Bradleys, or HMMWVs. Not only would this allow for a more dynamic and complex scenario, but it would provide valuable data on how group decisions are made and how UGVs may be utilized more than in the currently existing format. This information would be directly applicable to current design and development of UGVs for use within the next five to ten years.

D. CONCLUSION

The key finding from this study is that in a simple tactical scenario, tactical leaders tend to make the same decisions regardless of whether their wingmen are UGVs or live. A second key result is that military members are at least willing to utilize unmanned ground systems in certain combat situations. A great deal of technological growth and proving of technology will need to occur, however. The increased exposure and experience of military personnel with autonomous systems appears to have increased the level of trust in these systems, at least for very basic non-ambiguous, situations.

Some key actions must occur in order for future unmanned systems to be trusted to the extent that leaders will make unbiased decisions representative of this study. First, there must be an ability to have a human-in-the-loop. Even if a system eventually becomes completely reliable as a truly autonomous system, for trust to be possible there must be a way for a human to interact with the system to stop unexpected actions or thwart actions caused by a cyber-attack. Developers must continue to incrementally expose Soldiers and leaders to unmanned systems early (and often within their training), otherwise the unmanned systems will not be trusted in combat, if utilized at all. Lastly, there will need to be extensive studies and proofing of the technology to ensure that it will respond predictably and reliably in the most complex situations.

APPENDIX A. ADMINISTRATIVE GUIDE

SETUP Checklist:

- Generate Subject ID
  - Update subject List
- Input Subject ID into
  - initNew.sqf
  - Subject Data Tracker
  - Subject Survey Documents
- Update IP Address in initNew.sqf
- Update Subject Folder in Computer
- Conduct Trial run

Script:

Thank you for being part of the study. The purpose of this study is to gain insight into tactical level decision making. Specifically, we are also looking at the types of decisions people make when working with a live person, versus working with one that is automated. The scenario you will take part in is that of a Bradley Section leader. Depending on your study group, you will work with either a live wingman, or an automated wingman. Eye-tracking and EEG readings, combined with your decisions in the scenario and questionnaire, will allow us to better determine what sensory information is required, or can be used, to aide decision making. This will be used to help future development of Mission Command Systems as well as future Unmanned Systems.

Before I assign you to a study group I will need to have you fill out a consent form, sign in for your subject ID, provide some basic demographic information, and conduct some questionnaires. It is important for me to tell you that your information will be maintained in the strictest confidence. You will be assigned a subjectID that will be utilized in the data so that you cannot be linked to the information provided. Your personal information that identifies you will be maintained separately, in a secure location, and will only be used as a means to contact you during the course of the study, should we need to review any information or conduct any follow-ups. After the study has completed, your information will be disposed of within regulatory guidance to protect your identity and participation. This study is wholly voluntary and does contain some graphic battlefield scenes, if for any reason, you feel that you can no longer continue, or no longer want to participate, you are fully within your rights to stop and remove yourself from the study. None of your results or information will be used for the study and will be disposed in accordance with all regulatory guidance.

In front of you , you will find a subject packet which should correspond to your **subjectID, #________.** Please verify that your subjectID is correct at the top right of each page within the packet. During the course of the study, you will conduct two survey's and a small quiz. The first survey will gain some basic demographic information so that we can better judge what may have influenced your decision making. You will then conduct a basic quiz of tactical knowledge. Because we cannot create a scenario that matches the terms and methods of every service, the study was established using US Army doctrinal standards. As such, this quiz, will allow us to better understand how well you understand or relate to the terminology and techniques expressed within the study, mitigating any issues that may be caused by the use of unfamiliar terminology. Once complete, you will be fitted for the EEG and both the EEG and eye tracking systems will be calibrated to you.

At this time, please pull out the Pre-survey and Tactical Knowledge Quiz form your packet. All other documents can be placed back in the folder.   Fill out the survey and let the administrator, either myself or Mr Huston, know when you are complete.

<Fill Out the Survey and Quiz>

Now that you are complete with the surveys, Mr. Huston will fit you with the EEG and calibrate the systems. While he is doing that, I will use the information from your pre-survey and quiz to place you into your study group and initialize the study scenario.

We are using Virtual Battlefield Simulation 2 as our study medium, this is a networked first person serious-game, or game-for-training. You will use a mouse to interact with the screen to conduct any in game actions or decision making. Next to the first-person view you will also see a google map with tactical graphics. This map will show your location as well as your wingman. If your wingman sees an enemy target, they will attempt to populate the enemy on your map much like they would using an FBCB2 (Force XXI Battle Command Brigade and Below). We have provided a brief training scenario to allow you to become familiar with the program. I will start the training scenario for you now. When you are ready, please start the training scenario.

Okay, good job, do you feel that you understand the mechanics, or do you need to conduct the training scenario again?

You are going to be in the _________ group, your wingman will be _____________.

I have setup the mission for you and we are ready to begin. Please start at your convenience.

Very good job, you have now completed the experiment stage, we just need to complete another short questionnaire and conduct your debriefing and then we will be done.

Here is your questionnaire, I will be checking on the data, but please let me or Mr. Huston know when you are done, or if you have any questions.

<Fill out questionnaire>

Now that you are done, I can debrief you. You were part of the ___________ group.

<LIVE GROUP>

Your wingman was actually an autonomous agent. We attempted to deceive you in order to provide a more direct comparison to how people make decisions when using a live or autonomous system. To do this, we had to ensure that the live agent and the autonomous agent acted the same in both scenarios. While we could have used a live confederate and deceived the UGV group into believing they were working with a UGV versus a live person, this method ensured more commonality between scenarios and wingman actions. Please do not impart this knowledge to anyone else that would be a possible candidate for this study as having this knowledge previously would seriously jeopardize the validity of the study. If you do not agree with our deception, you may opt to withdraw from the study, and we can omit all data we collected from you. Would you like to opt out of the study?

Would you like to know how you did in your decision making?

Yes – The data we have so far, is very preliminary and there is no necessarily right or wrong answers in how you conduct this scenario; but we did evaluate all the choices to provide a scoring criteria of what paths stayed the closest to US Army doctrine. Like golf, a low score is better, whereas a high score (the max is 9) indicates a less desired path choice. If you would like the final results of the study, I will send the completed thesis to you after it has been approved by the Thesis department.

<UGV GROUP>

Your wingman was actually an autonomous agent, but so was the live group. We attempted to deceive those individuals in the live group in order to provide a more direct comparison to how people make decisions when using a live or autonomous system. To do this, we had to ensure that the live agent and the autonomous agent acted the same in both scenarios. While we could have used a live confederate and deceived the UGV group into believing they were working with a UGV versus a live person, this method ensured more commonality between scenarios and wingman actions. Please do not impart this knowledge to anyone else that would be a possible candidate for this study as having this knowledge previously would seriously jeopardize the validity of the study. If you do not agree with our deception, you may opt to withdraw from the study, and we can omit all data we collected from you. Would you like to opt out of the study?

Would you like to know how you did in your decision making?

Yes – The data we have so far, is very preliminary and there is no necessarily right or wrong answers in how you conduct this scenario; but we did evaluate all the choices to provide a scoring criteria of what paths stayed the closest to US Army doctrine. Like golf, a low score is better, whereas a high score (the max is 9) indicates a less desired path choice. If you would like the final results of the study, I will send the completed thesis to you after it has been approved by the Thesis department.

Follow-On Tasks:

Open scenario log files and transfer information into study Excel sheet.

Clear SubjectID information from .sqf files to prepare for next subject.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. DEMOGRAPHIC SURVEY

Subject ID #: ______________

What United States Military Service are you?

US Army   US Navy   US Marine Corps   US Air Force

What is your Job Category you most associate with based on your entire military service?

Combat Arms   Special Forces   Service Support   Combat Service Support

Are you male or female?

Male   Female

What is your Age? ___________

Years of Military Service? __________

How many combat tours have you been on? This applies to Ground Combat Experience into hostile areas such as Iraq, Afghanistan, Boznia, etc.

Average Length of Combat Tours (months)?

What kind of Mounted Vehicle Experience do you have?

Tracked Vehicles   Wheeled Vehicles   Both   None

What percentage of your military service has been mounted? 0 – 100%   ________

Were you trained in mounted Combat?   Yes   No

Have you ever been in direct contact with an enemy force while in a mounted vehicle and as a leader?

Yes   Yes, but not as a leader   No

Was your experience in a conventional fight or in low intensity conflict (COIN, counter-insurgency)?

Conventional   Low Intensity   Both   N/A

Have you had any experience with unmanned vehicles? Circle the response that best represents your experience.

No   Used   Operated   Directed   Used and Operated

Used and Directed   Operated and Directed   Used, Operated, and Directed

What percentage of your military service have you used, operated, or directed Unmanned Vehicles?

0 – 100% ________

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. TACTICAL KNOWLEDGE TEST

Subject ID #: _______________

Pre-Test of Tactical Knowledge

1. What is a Movement To Contact? (Circle one)

A. Movement to contact is a type of defensive operation designed to develop the situation and establish or regain contact.

B. Movement to contact is a type of offensive operation designed to attack known enemy positions.

C. Movement to contact is a movement technique designed to allow freedom of movement without contact.

D. Movement to contact is a type of offensive operation designed to develop the situation and establish or regain contact.

2. Of the Three Movement Techniques, Travelling, Travelling-Overwatch, Bounding-Overwatch; what are the main differences and when are they used? (Circle One)

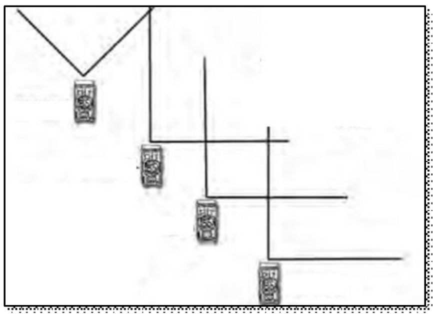
A. Traveling – used when speed is necessary and contact with enemy forces is not likely.
Traveling Overwatch – movement technique when contact with enemy forces is possible, but speed is important. The lead element is continuously moving while the trailing elements move at variable speeds, sometimes pausing to overwatch movement of the lead element.
Bounding Overwatch - movement technique when contact is expected with enemy forces. One type of bounding overwatch consisting of alternating bounds

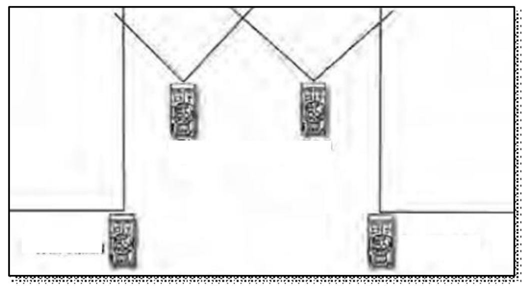
B. **Traveling** – used when speed is not necessary and contact with the enemy is expected.
Traveling Overwatch – movement technique when contact with enemy forces is possible, but speed is important. The lead element is continuously moving while the trailing elements move at variable speeds, sometimes pausing to overwatch movement of the lead element.
Bounding Overwatch – movement technique when contact is expected with enemy forces. Two types of bounding overwatch: alternating bounds and successive bounds.

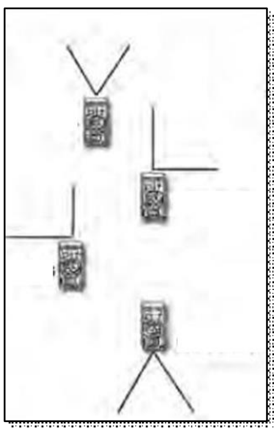
C. Traveling – used when speed is necessary and contact with enemy forces is not likely.
Traveling Overwatch – movement technique when contact with enemy forces is possible, but speed is important. The lead element is continuously moving while the trailing elements move at variable speeds, sometimes pausing to overwatch movement of the lead element.
Bounding Overwatch - movement technique when contact is expected with enemy forces. Two types of bounding overwatch: alternating bounds and successive bounds.

D. Traveling – used when speed is necessary and contact with the enemy is not likely.
Traveling Overwatch – movement technique when contact with enemy forces is possible, but speed is important. The lead element is continuously moving while the trailing elements move at variable speeds, sometimes pausing to overwatch movement of the lead element.
Bounding Overwatch - movement technique when contact is not expected with enemy forces. Two types of bounding overwatch: leaping bounds and successive bounds.

3. Below are several pictures of Movement Formations. Please indicate where the Platoon Leader would be located in the formations. If there is a difference to their location based on movement technique, please list both locations and annotate in parenthesis what movement technique they would be in for that location(Circle the vehide where you think the PL belongs)







APPENDIX D. POST SURVEY

Subject ID #: _______________

Post Survey

All answers are on a scale from 1 – 5, with five being the most confident and 1 being the least confidence.

MD1. Your decision was__________.

Why did you make choose this decision?

____

____

____

TD1. Your decision was__________.

Why did you make choose this decision?

____

____

____

TD1a. Did you request more time?

Yes - # times =

No

Why?

____

____

____

MD2. Your decision was__________.

Why did you make choose this decision?

____

____

____

TD2. Your decision was__________.

Why did you make choose this decision?

____

____

____

TD2a. Did you request more time?

Yes - # times =

No

Why?

____

____

____

Would you trust an autonomous System? Why or Why not? By autonomous we mean a system that is self-reliant, with no operator directly controlling the system, but which receives guidance from a leader similar to a live wingman or as depicted in the scenario.

____

____

____

____

Would you like a copy of the study in its final form?

Yes No

Now that you have completed the task, what overall level of confidence do you have in your decisions?

0 – 100%

What was your level of trust in your wingman?

0 – 100%

What aspects did you most trust?

Accuracy of Information – the information they provided was timely and correct.

Expected Actions – able to execute the task provided in your decision

Knowledge of Situation – the wingman knew enough about what was going on around them to give adequate information or execute tasks effectively.

What aspects did you least trust?

Accuracy of Information – the information they provided was timely and correct.

Expected Actions – able to execute the task provided in your decision

Knowledge of Situation – the wingman knew enough about what was going on around them to give adequate information or execute tasks effectively.

APPENDIX E. APPROVED IRB PROTOCOL

Naval Postgraduate School
Human Research Protection Program

DEC 23 2013

From: President, Naval Postgraduate School (NPS)
To: Dr. Quinn Kennedy, Operation Research Department (OR)
Mr. Jesse Huston, Modeling, Virtual Environments and Simulation Institute (MOVES)
LTC Jonathan Alt, USA
MAJ Peter Nesbitt, USA
MAJ Scott Patton, USA
Via: Chairman, Institutional Review Board (IRB)

Subj: A COMPARISON OF TACTICAL LEADER DECISION MAKING WITH AUTOMATED OR LIVE COUNTERPARTS IN A VIRTUAL ENVIRONMENT (VIRTUAL BATTLEFIELD SIMULATION 2)

Encl: (1) Approved IRB Protocol

1. The NPS IRB is pleased to inform you that the NPS President has approved your project (NPS IRB# NPS.2014.0009-IR-EP7-A). The approved IRB Protocol is found in enclosure (1). Completion of the CITI Research Ethics Training has been confirmed.

2. This approval expires on 30 June 2014. If additional time is required to complete the research, a continuing review report must be approved by the IRB and NPS President prior to the expiration of approval. At expiration all research (subject recruitment, data collection, analysis of data containing PII) must cease.

3. You are required to obtain consent according to the procedure provided in the approved protocol.

4. You are required to report to the IRB any unanticipated problems or serious adverse events to the NPS IRB within 24 hours of the occurrence.

5. Any proposed changes in IRB approved research must be reviewed and approved by the NPS IRB and NPS President prior to implementation except where necessary to eliminate apparent immediate hazards to research participants and subjects.

6. As the Principal Investigator (PI) it is your responsibility to ensure that the research and the actions of all project personnel involved in conducting this study will conform with the IRB approved protocol and IRB requirements/policies.

Subj: A COMPARISON OF TACTICAL LEADER DECISION MAKING WITH AUTOMATED OR LIVE COUNTERPARTS IN A VIRTUAL ENVIRONMENT (VIRTUAL BATTLEFIELD SIMULATION 2)

7. At completion of the research, no later than expiration of approval, the PI will close the protocol by submitting an End of Experiment Report.

Lawrence G. Shattuck, PhD
Chair
Institutional Review Board

Ronald A. Route
Vice Admiral, U.S. Navy (Ret.)
President, Naval Postgraduate School

Date: DEC 23 2013

2

APPENDIX F. INFORMED CONSENT

Naval Postgraduate School
Consent to Participate in Research

Introduction. You are invited to participate in a research study entitled *"A Comparison of Tactical Leade Decision Making with Automated or Live Counterparts in a Virtual Environment."* The purpose of the research is to determine if a person's decisions in tactical situations change based on the use of a live wingman versus an automated wingman.

The scenarios used in this study provide a realistic and graphical depiction of combat. If you have been diagnosed with PTSD or feel that you may have PTSD like symptoms you should not participate. By consenting to participate below you acknowledge that you are aware of the risks associated with your participation.

Procedures.

- You will first receive a brief overview of the study and your involvement. Following the brief, you will fill out a standard consent form, providing both written and oral consent to al or part, of the experiment. After you have filled out the consent form, you will logged into the subject log and receive your subject ID number. Both the subject log and the consent form will be maintained in hard copy document only. These documents are the only place i which your name and contact information will appear. The log and consent form will be stored separate from data within a secure location in the Primary Investigator's office. All other data will be stored electronically on access controlled computers at the Naval Postgraduate School. Data from the experiment will only be referenced using the subject ID number. All data in the final report will be reported in aggregate.
- You will fill out a pre-survey questionnaire, take a five-minute test to determine your baseline tactical experience, and fill out a demographic survey.
- You will be required to be go through the setup process for the EEG and eye tracking devices. This data will help us understand how leaders may make decisions in a tactical environment. Your name will not be on any of the data used by the study, only your subject ID number.
- You will then conduct a familiarization scenario to ensure you understand the controls and capabilities within the system. You will conduct one of two scenarios (Live or UGV) in which they will act as the Bradley Section Leader in an Army Game for Training program known as Virtual Battlefield Simulation 2 (VBS2). At the completion of the scenario, you will be given a post-experiment survey capturing data necessary to better understand the scenario results. Lastly, you will be debriefed about the experiment, ensuring that you understand all aspects of the experiment. In total, the time for one trial should not exceed 2 hours. Between 30-40 other subjects will participate in this research.
- Data will be reported in aggregate; though specific results may be referenced in academic publications with no reference to you, the subject. Your individual results will not be released to any entity or agency and you will never be identified in any publications.
- There will be no audio or video recording, though the scenario will be recorded using a built-in Afte Action Review tool, in order to synchronize specific data points within the study between the virtual environment, the EEG, and the eye-tracking device.

Location. The interview/survey/experiment will take place at the Naval Postgraduate School, in Watkins Hall.

as a research subject or any other concerns may be addressed to the Navy Postgraduate School IRB Chair, Dr. Larry Shattuck, 831-656-2473, lgshattu@nps.edu.

Statement of Consent. I have read the information provided above. I have been given the opportunity to ask questions and all the questions have been answered to my satisfaction. I have been provided a copy of this form for my records and I agree to participate in this study. I understand that by agreeing to participate in this research and signing this form, I do not waive any of my legal rights.

_____________________________________ Date _________________

Participant's Signature

_____________________________________ Date _________________

Researcher's Signature

APPENDIX G. MISSION BRIEF

Thank you for partaking in the study.

Task:
Your task is to act as the Section leader for a Bradley Section that is conducting a Situational Tactical Exercise (STX). You, and your wingman will maneuver through the STX lane based on the decisions you make in the dialogue windows.

Situation:
A dismounted Infantry Squad, consisting of 9 dismounts, is believed to be moving from North to South through your AO. They are part of a larger Yurogazian force comprised of two Infantry Company's and one Mechanized Reconnaissance Company. The Yurogazian's have very few actual mechanized vehicles and generally will stage them forward, relying on the speed of their wheeled vehicles to get them back to their defenses after identifying any enemy threat.

Mission:
Conduct a movement to contact South to North from Phase Line Dakota (LD) to the LOA in order to identify and/or destroy any enemy threat moving through the region. The commander's intent is that the region is free of any enemy forces.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX H. LEVEL OF EXPERIENCE METRIC

A participant's experience level only determines the subjects experience level as it relates to the study. Therefore, a person with a considerable number of years of service, but no mounted experience or unmanned vehicle experience, may still be classified as middle or low experience.

A. Demographic weighting parameters

To determine the experience level of each subject, their demographic data they provided and their tactical knowledge tests were combined to provide a score ranging from 0 to 412. The levels of experience were broken down so that high scores range from 412 to 285. Middle scores range from 284 to 155. Low scores range from 156 to 0. In order to calculate a score, the various possible answers provided by the demographic survey were weighted out of a range from 0 to 100 based on the following distribution:

Weighting Scale	
Years of Service	20
Mounted Exp	5
% Mounted Time	13
Trained?	5
UAV Exp	5
%UAV Exp	12
Combat Tours	0
Avg length of Tours	5
Direct Contact?	3
LIC/Conv.	5
JOB	5
Service	2
Tactical Knowledge Test	20

Table 11. Experience metric weighting scale

All scores were evaluated individually and provided a score ranging from 0 to a max, in one case, of 7. Some demographic answers had to be combined to provide more relative responses and quantifiable results. Figure 32 shows the evaluation metrics are for each category along with any calculations performed.

Experience Scores for Demographic Characteristics

Service	
Answer	**Score**
Army/Marine	3
USAF	2
USN	1

Years of Service	
Answer	**Score**
>=10	4
>=7 && < 10	3
>=4 && < 7	2
>=1 && <4	1
<1	0

Job Classification	
Answer	**Score**
Combat Arms / Special Forces	3
Combat Service Support	2
Service Support	1

Combat Tours	
= # combat tours * Average tour length	
Answer	**Score**
>=36	5
>=24 && <36	4
>=12 && <24	3
>=6 && <12	2
>=1 && <6	1
0	0

Direct Contact	
Answer	**Score**
Yes - as leader	2
Yes - not as leader	1
No	0

Type of Conflict Experience	
Answer	**Score**
Both	3
Conventional	2
Low Intensity Conflict	1
None	0

Mntd Exp And Yrs of Service	
= # years service * % Mounted Experience	
Answer	**Score**
>= 8	6
>=6 && <8	5
>=4 && <6	4
>=2 && <4	3
>=1 &&<2	1
>0 && <1	0.5
0	0

Mounted Experience	
Answer	**Score**
Both	3
Tracked	2
Wheeled	1
None	0

Trained in Mounted Warfare	
Answer	**Score**
Yes	1
No	0

Unmanned Systems Experience	
Answer	**Score**
Used,Operate,Direct	7
Used and Directed	6
Used and Operated	5
Operated and Directed	4
Directed	3
Operated	2
Used	1
None	0

% of Unmanned Experience	
Answer	**Score**
>= 80%	6
>=60% && <80%	5
>=40% && <60%	4
>=20% && <40%	3
>=10% &&<20%	2
>0% && <10%	1
0%	0

Figure 32. Experience scores for demographic characteristics

The resulting scores per demographic category were then summed based on the weighting scheme to provide a general idea of how experienced the participant was in the critical aspects of this study. The final experience rating was determined using the following formula:

$$\begin{aligned} \text{Experience Rating} = {} & \text{Tactical Knowledge Test} * 20 \\ & + \text{Years of Service} * 20 \\ & + \text{Mounted Experience} * 5 \\ & + \text{Percentage of Time in Service Mounted} * 13 \\ & + \text{Trained in Mounted Warfare} * 5 \\ & + \text{Type of Unmanned Vehicle Experience} * 5 \\ & + \text{Percentage of Time with Unmanned Systems} * 12 \\ & + \text{Combat Tours} * 5 \end{aligned}$$

Equation used to calculate participant experience level from provided demographic data

The three experience levels, “High”, “Mid”, and “Low” were determined to be relatively equally distributed but with a slightly larger range for low level of experience in order to ensure key attributes (such as mounted or unmanned system experience) had to be present to achieve mid to high level experience ratings.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX I. DETAILED DEPICTION OF HUD

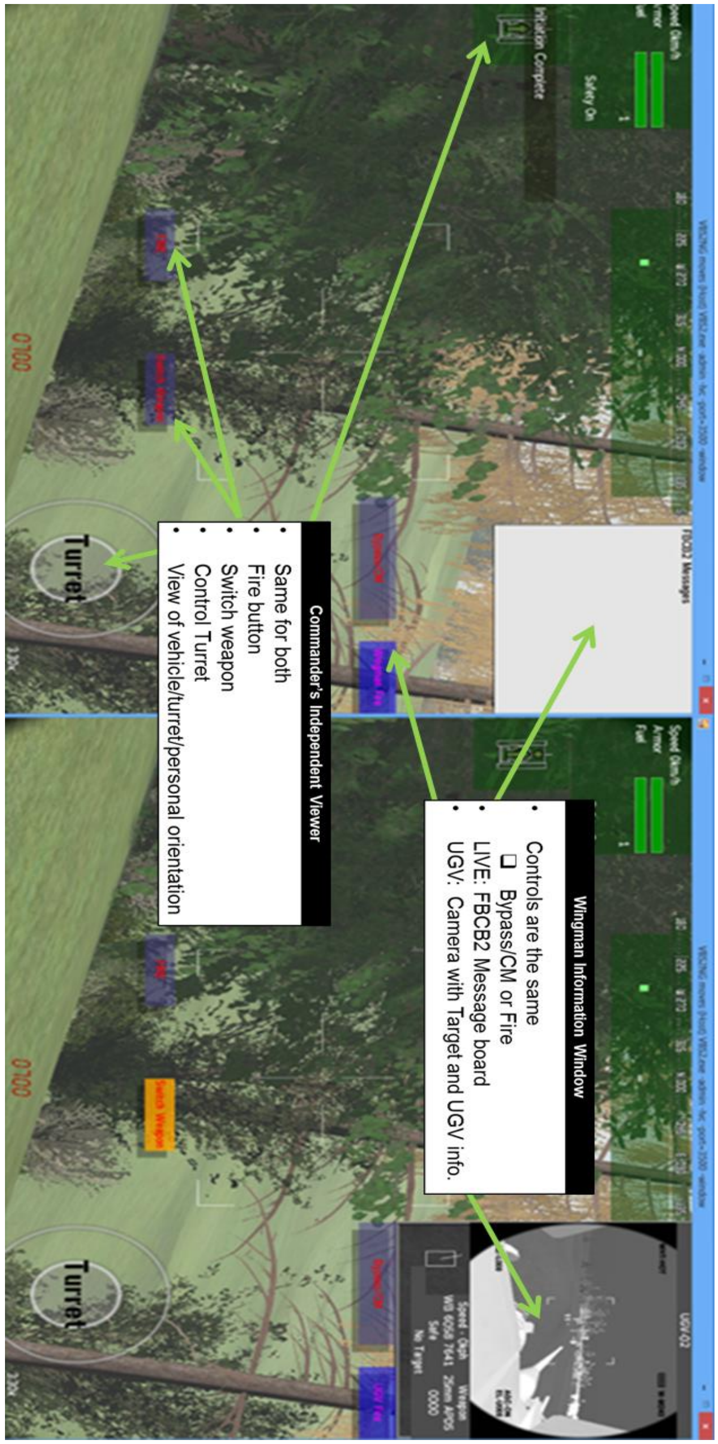


Figure 33. Live (left) and UGV (right) HUD depictions with explanation of characteristics

THIS PAGE INTENTIONALLY LEFT BLANK

SUPPLEMENTAL A. VIRTUAL BATTLESPACE 2 SCRIPTS

Script files for the study are too large to place in a single document. VBS2 creates a set of files whenever a scenario is created using the simulations mission planning editor. These files are stored in “MPMissions,” within the VBS2 folder that is stored in the computer’s “Documents” folder. The supplemental contains the entire set of VBS2 code, both that created by VBS2 and that created by the design team, to allow for a person with reasonable knowledge of VBS2 to conduct the experiment. A flow chart is included in Figure 32 to provide a better understanding of how the primary files are utilized.

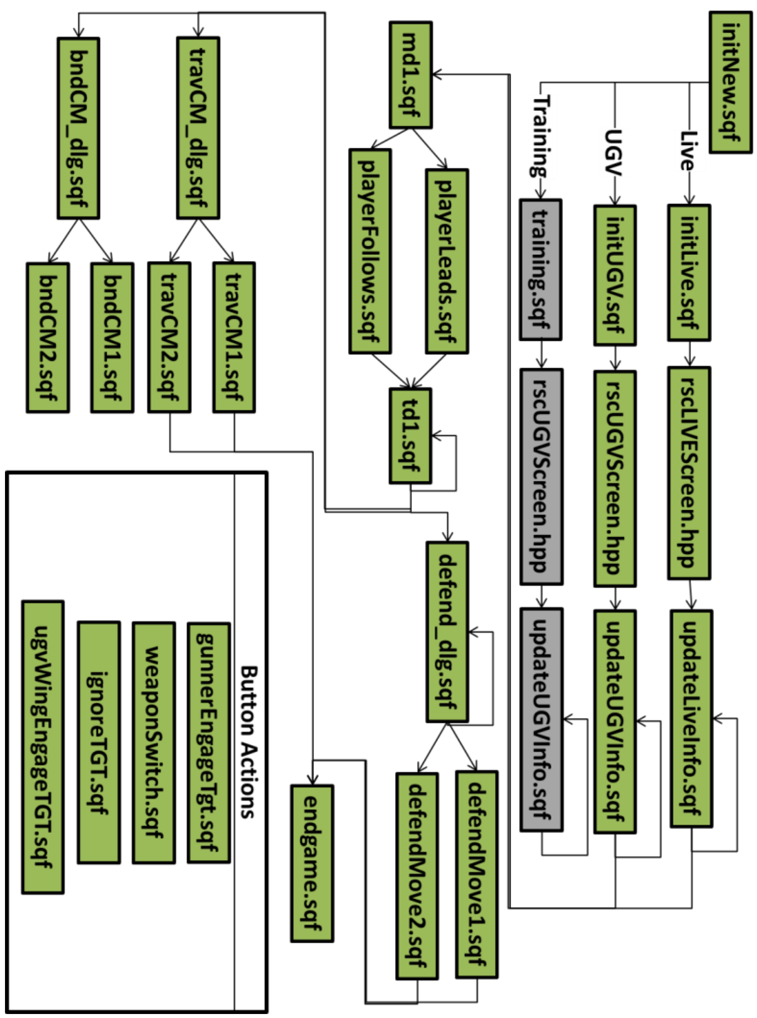


FIGURE 34. Primary VBS2 script flow

THIS PAGE INTENTIONALLY LEFT BLANK

SUPPLEMENTAL B. WEBSOCKET GATEWAY

The websocket gateway was created by Professor Donald McGregor, a Research Associate at the Naval Postgraduate school. He modified the code to allow for multithreading and the project team modified the index.htm file to allow for google maps to represent a 2D 1:50,000 map display of unit movements within the simulation. The file was coded using Netbeans and is a mixture of Java and Javascript. The code is open source, but is intended for limited academic use in order to stay in compliance with google licensing practices which allow for the use of their google maps protocols.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Bohemia Interactive Simulations. (2014, 04 23). *BISim Wiki*. Retrieved from bisimulations.com: https://resources.bisimulations.com/w/index.php?title=Main_Page

Colebank, J. L. (2008). *Automated intelligent agents are they trusted members of military teams?* (Master's thesis, Naval Postgraduate School). Retrieved from http://hdl.handle.net/10945/3868.

Cruz, A. S. (2011, June). The robot general. *Armed Forces Journal*. Retrieved July 25, 2013, from www.armedforcesjournal.com/2011/06/6187209/

Hancock, P. A., Billings, D. R., Oleson, K. E., Chen, J. Y., De Visser, E., & Raja Parasuraman, R. (2011.). *A meta-analysis of factors influencing the development of human-robot trust*. Retrieved August 05, 2013, from http://www.dtic.mil/dtic/tr/fulltext/u2/a556734.pdf

Klein, G. A., & Calderwood, R. (1990). *Investigations of naturalistic decision making and the recognition-primed decision model*. Ft. Belvoir, VA: Defense Technical Information Center.

Kobus, D. A., Kobus, J. M., Ostertag, J., Kelly, M., & Palmer, E. D. (2010). *Assessing changes in tactical decision making: FITE-JCTD spiral 2 operational demonstration*. Arlington, VA: Pacific Science & Engineering Group.

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50–80. doi: 10.1518/hfes.46.1.50_30392

Legree, P. J., & Psotka, J. (2006). *Refining Situational Judgment test methods*. Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

McCaney, K. (2014, April 10). *Virtual reality gets real with Army's 3D game update*—Defense Systems. Retrieved from http://defensesystems.com/articles/2014/04/10/army-vbs3-virtual-training-gets-real.aspx

Milgaming. (2010). *VBS2*. Retrieved from https://milgaming.army.mil/VBS2/default.aspx

OC, I. (2011). *Department of Defense dictionary of military and associated terms*. [Washington, DC]: United States Joint Chiefs of Staff.

Robotic Systems, Joint Project Office. (2011). *Unmanned Ground Vehicle (UGV) Interoperability Profile (IOP)*. Warren, MI: US Army.

Trust. (2014) In *Merriam-Webster Online Dictionary*. Retrieved from http://www.merriam-webster.com

United States. Defense Science Board. Task Force on the Role of Autonomy in DOD Systems., & United States. Office of the Under Secretary of Defense for Acquisition, T. (2012). *Task force report: The role of autonomy in DOD systems*. Washington, DC: Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics.

United States. (2001). *Tactics: Field manual 3–90*, Washington DC Headquarters, Dept. of the Army. Retrieved from https://www.armypubs.army.mil.

United States. (2012, May). *The operations process: Army Doctrinal Publication 5–0*. Washington DC: Headquarters, Department of the Army. Retrieved from https://www.armypubs.army.mil.

United States. (2012). *The operations process: Army Doctrinal Reference Publication 5–0*, Washington DC: Headquarters, Department of the Army.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California