Operator Suspicion and Detection / Response to Cyber-Attacks on Unmanned Systems

---

A Dissertation

Presented to the Faculty of the School of Engineering and Applied Science

The University of Virginia

In partial fulfillment

Of the requirements for the degree of

Doctor of Philosophy in Systems Engineering

Christopher A. Gay

May 2017

Approval

This dissertation is submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Systems Engineering

_____

Christopher A. Gay, Author

The dissertation has been read and approved by the examining committee:

_____

William Scherer, Committee Chair

_____

Barry Horowitz, Co-advisor

_____

Inki Kim, Co-advisor

_____

Michael Smith, Committee Member

_____

Eileen Chou, Committee Member

_____

John Elshaw, Committee Member

Accepted for the School of Engineering and Applied Science:

_____

Dean, School of Engineering and Applied Science

May 2017

Abstract

Cyber-attacks against cyber-physical systems, such as unmanned vehicle systems, are serious and emergent threats with potentially catastrophic impacts, and the topic has garnered considerable interest. Much research is being done to address the physical security aspects of cyber-physical systems; however, research addressing the human dimension of cyber-attack detection and response from an operator and operational perspective is sparse. My research was a novel probe into the human factors affecting operator resilience to cyber-attacks, which are situations characterized by uncertainty and malicious intent. The variability of individual operators makes it improbable to grasp the full range of factors contributing to operator performance; however, the literature review provided a starting point to aid in understanding operator performance in situations involving malicious intent (e.g. a cyber-attack). Malicious intent is a component of the suspicion theory developed by (P. Bobko, Barelka, & Hirshfield, 2014), and suspicion was believed to be a key factor in operator response to cyber-attacks. The research effort explored this human dimension through scenario based, human-in-the loop behavioral science experiments with Air Force personnel. It included both abstract and empirical assessments of the application of suspicion theory to operator detection and response to cyber-attacks against an unmanned vehicle system, and it took a systems-oriented approach to the problem by incorporating a human-machine team (HMT) in the response. The HMT was defined as an operator (human) and a *Sentinel* (an automated hardware / software cyber-attack detection aid). The study allowed for the a) evaluation of the relationship between general trait-level attributes and operator suspicion, b) analysis of the effects of suspicion on the operator and *Sentinel* team performance, and c) study of the effects of consequences and perception of consequences on operator suspicion and performance.

Acknowledgments

First and foremost, I give thanks to my Lord and Savior Jesus Christ for strengthening and sustaining me throughout this arduous doctoral process. I'm also thankful for the many Christian brothers and sisters who supported me in prayer and encouraged me along the journey.

I am grateful to my wife, Sharon, and my children, Jason and Rachel, for the love, support, and understanding they've provided throughout this academic pursuit and the 19+ years of military service. I know it has not been easy (or even fair) for you at times, but you support me no matter what. Your unconditional love and support are the reasons for my success in all of life's endeavors. You are my rock.

I am grateful to my mother, Nancy Coleman, and my brother, Steve Gay, and his family for their prayers, love, and support throughout this journey and all of my life journeys.

I would like to express my gratitude to my co-advisors, Professor Barry Horowitz and Professor Inki Kim, for their guidance and support throughout my doctoral program. I am grateful to my dissertation committee: Professor William Scherer and Professor Michael Smith of the Department of Systems and Information Engineering, Professor Eileen Chou of the Batten School, and Professor John Elshaw of the Air Force Institute of Technology, for their invaluable knowledge and assistance.

I am grateful to Dr. Phil Bobko for his guidance, encouragement, and friendship throughout my doctoral research endeavors. I'm also thankful for the support of Mr. Jeremy Gray and Dr. Svyatoslav Guznov for their assistance in operationalizing the experiment.

I am grateful to my leadership and co-workers at the Air Force Institute of Technology for sponsoring me into this doctoral program and being steadfast in their support.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Chapter Overview

This chapter introduces the topic of the dissertation. It describes the motivation for the topic and provides an outline for the organization of the dissertation.

1.2 Motivation

When considering the operation of a cyber-physical system, such as a remotely piloted aircraft system (RPAS) in a cyber contested environment, two broad categories of variables impact overall system performance and resilience to cyber-attacks: physical (hardware / software) system performance and operator (human) performance. Multiple variables affect performance in each of these categories. The University of Virginia's (UVA) system aware cyber-security (Jones & Horowitz, 2012a) (Horowitz & Pierce, 2013) and *Sentinel* (Horowitz & Jones, 2015)(Gay et al., 2015) research are examples of work being done to study the performance of physical systems in this context; however, research addressing the human dimension of cyber-attack response from an operator and operational perspective is sparse. These systems represent an intrinsic vulnerability for adversaries to perform cyber-attacks for counter-control or subversion of military assets. As an example, Iranian cyber capabilities were believed to have brought down the Central Intelligence Agency operated RQ-170 Sentinel drone operating near the Iranian border. The Iranians successfully landed the drone in December 2011, causing grave concern over potential compromise of highly sensitive surveillance capabilities. This is an emergent area of research due to the potential devastation that can result from cyber-attacks against cyber-physical systems. Unlike information technology systems, operators of

cyber-physical systems must respond to cyber-attacks in real-time to prevent potentially catastrophic loss of the physical system (e.g. RPAS) and its highly classified components or the unintentional loss of human life should the cyber-attack divert the firing of the weapons system or cause it to malfunction.

The DoD System Engineering Research Council (DoD SERC) funded UVA to explore the development of a cyber-security concept of operations (CONOPS) for the Air Force RPAS. UVA, in a partnership with MITRE Corporation and Creek AFB, performed this study over the 2013-2014 timeframe. The UVA / MITRE team presented a report to the DoD SERC on the findings from the 2013-2014 study effort (Gay et al., 2015). The following were some of the findings that motivated my research effort: 1) Operators unwittingly subjected to cyber-attacks were unable to detect them without the assistance of a *Sentinel* automated cyber-attack detection aid, 2) Operators did not consider the issues cyber-attacks and eventually aborted some of their missions and returned to base after exhausting their normal maintenance and operations checklists, 3) When the *Sentinel* aid was present and alerted the operator of a cyber-attack, the operators were unsure how to respond to the *Sentinel* alerts, and 4) The operators did not suspect malicious intent.

Given the sophistication and potential consequences of cyber-attacks targeting RPAS missions to thwart military operations (e.g. RQ-170 Iranian "incident"), there is a great need for a *Sentinel*-like capability to aid the operator in detecting cyber-attacks. However, a technology solution alone may not provide the level of security required, since technology itself is fallible as it may not be configured to detect the latest emergent threat, or it could generate an alert to anomalous system behavior unrelated to a cyber-attack, such as a maintenance issue. Ultimately, the human operator is the decision maker. In all cases the operator must determine whether the

Sentinel aid is correct or not from all of the available information. In some situations, the operator could conceivably need to override the *Sentinel* and in other cases, the operator may need to intervene when the *Sentinel* does not. Therefore, performance against a cyber-attack must be viewed from a systems-oriented perspective of a human-machine team (e.g. an operator and *Sentinel* team, HMT) with emphasis on the operator's ability to accurately assess and respond to a given situation.

Motivated by these findings, this research probes into the factors affecting operator resilience to cyber-attacks, which are situations characterized by uncertainty and malicious intent. As with physical system performance, many potential variables (e.g. emotion, trust, cognitive ability, creativity, situational awareness, etc.) contribute to operator performance. The variability of individual operators makes it improbable to grasp the full range of factors contributing to operator performance in every situation. Fortunately, the literature provides a starting point to aid in understanding operator performance in situations involving malicious intent (i.e. a cyber-attack). The theory of suspicion proposed by Bobko, Barelka, and Hirshfield (P. Bobko et al., 2014) offers a "lens" through which to view the critical issue of operator response to cyber-attacks; they wrote, "It is the simultaneous combination of uncertainty, perceived malintent, and increased cognitive activity that defines state suspicion," and they made several propositions regarding the utility of the theory. Bobko et al.'s concept of state suspicion appeared related to the findings from the earlier 2013-2014 UVA / MITRE RPAS DoD SERC effort, and it has potential to influence operator response to cyber-attacks.

Thus, the primary goal of this research is to study the relationship of operator suspicion to the detection and response of cyber-attacks in a human-machine team (HMT) context through the application of suspicion theory to scenario base, human-in-the loop behavioral science

experiments involving operators of a representative cyber-physical system in a cyber-contested environment. The experiments manipulated the dimensions of suspicion through a range of scenarios, measured operator suspicion in each situation, and determined HMT performance outcomes for each case. The data collection and analysis provided novel empirical evidence of the utility of suspicion theory in operator detection and response to cyber-attacks. The results of the analysis will be presented to the Air Force and DoD personnel and provided to the suspicion research community to further advance research in this area as shown in **Figure 1**. **Figure 1** is a top-level diagram of the research effort. The light blue box in the center of diagram titled, "Operator Cyber Response Research," was the primary focus of this study, and the objective was to address the following questions within the context of a cyber-contested environment:

1) What is the relationship between suspicions and HMT performance?

2) How does consequence effect the relationship between suspicion and HMT performance?

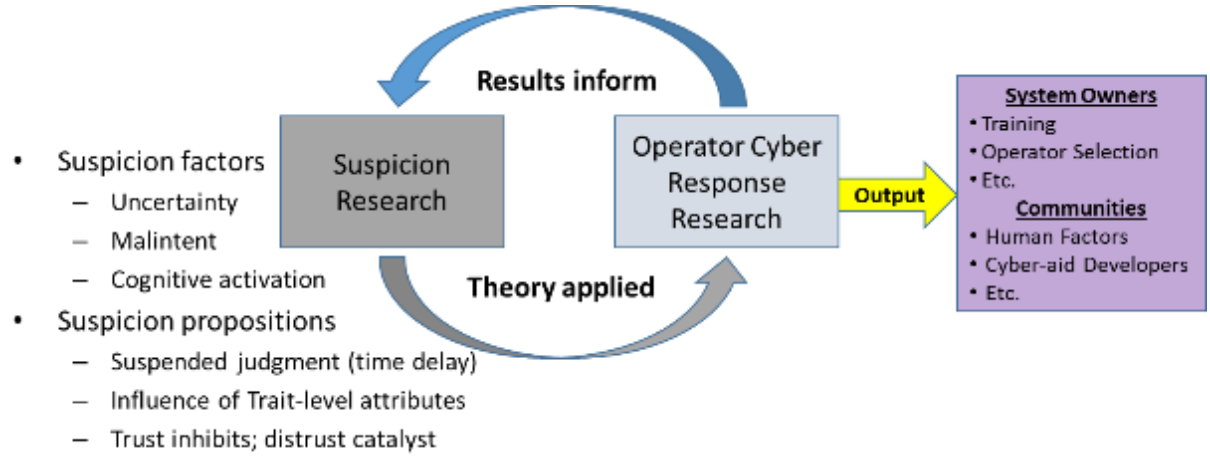


Figure 1: Diagram of Research Effort

1.3 Organization of the Dissertation

The dissertation is organized as follows. Chapter 2 presents the current literature on suspicion theory and its application. The discussion of this literature forms the foundation for

the research. It also discusses the research questions and hypotheses associated with this research effort. Chapter 3 presents the methodology and design of experiment (DOE) necessary to operationalize the theory of suspicion in a statistically relevant way to address the research questions and hypotheses. Chapter 4 offers a discussion of the questions and hypotheses presented in Chapter 2 in light of the data and results from the experiments and presents the key findings, concerns and limitations. Chapter 5 presents the summary and conclusions of the research and discusses research contributions and future work.

Chapter 2: Theoretical Model and Hypotheses

2.1 Chapter Overview

This chapter introduces the concept and theory of suspicion, which forms the foundation for the research. The chapter begins with a review of the literature and links the theory of suspicion to its application enabling the design of experiments to address the key questions and hypotheses associated with the research.

2.2 Construct of Suspicion Theory

A review of current literature in the domains of trust and suspicion determined the theory of suspicion – as proposed by Bobko et al. (P. Bobko et al., 2014) – to be the most relevant literature for a study of operator response to cyber-attack due to its emphasis on perception of malicious intent and its focus on information technology (IT) related contexts. This theory of state-suspicion was developed under a research effort sponsored by the Air Force Research Lab (AFRL) and led by the 711th Human Performance Wing (711HPW). It was funded by the Air Force Office of Scientific Research (AFOSR), and it spanned multiple social science domains

including psychology, human factors, marketing, management, and communication. Collaboration continued with Dr. Bobko and AFRL to ensure the appropriate application of the suspicion theory and one or more models for studying operator response to cyber-attacks. The theoretical definition of state-suspicion proposed by Dr. Bobko (P. Bobko et al., 2014) for an information technology (IT) related context was:

> "State suspicion is a person's simultaneous state of cognitive activity, uncertainty, and perceived malintent about underlying information that is being electronically generated, collated, sent, analyzed, or implemented by an external agent."

2.3 State-Suspicion Model

In this section, I present the State-Suspicion model developed by Bobko et al. shown in **Figure 2** and give a brief explanation of the theory. All references to Stage-levels refer to the model in **Figure 2**. Section 2.4 includes a discussion of research propositions taken from the literature review associated with the suspicion model and applied to this study.

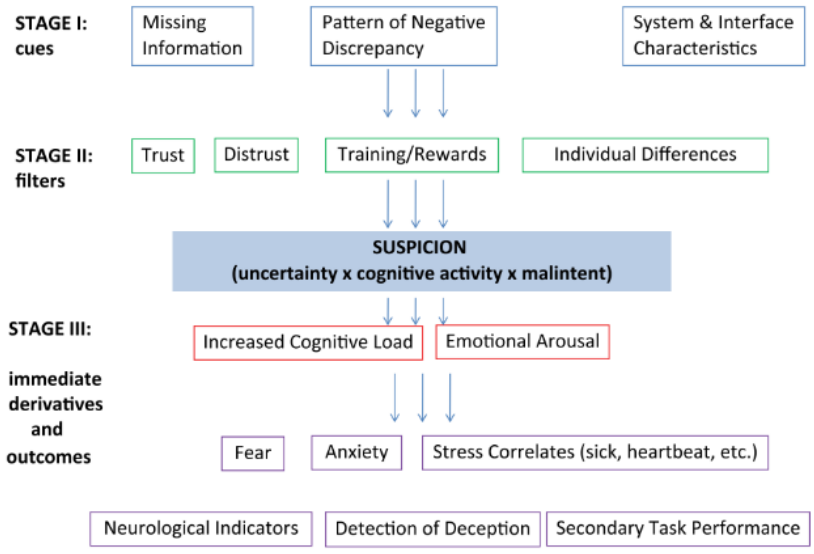


Figure 2: Stages of State-level IT Suspicion (P. Bobko et al., 2014)

Stage I cues referred to indications from the environment which can act as a trigger for state-level suspicion. The boxes listed across the "Stage I: cues" row were examples of categories of potential indicators that can serve as sources of manipulation for the experiment. Listings of more specific prompts than those shown in **Figure 2** were included in their article. The test construct for this research used an operator and *Sentinel* pair as the human-machine team (HMT) for detection of cyber-attacks. The *Sentinel* alerts served as environmental cues to the operator for manipulation during the study.

Stage II filters denoted individual difference (trait-level) variables likely to affect state-level suspicion. In this model trust and distrust refer to an individual's propensity towards those factors. Schoorman, et al. included a seven item measure for propensity to trust in their 2007 article and some researchers have reverse scored the measures to account for distrust (Schoorman, Mayer, & Davis, 2007). The propensity to trust or distrust was interesting from a behavioral science point of view, because they both potentially affect suspicion. Suspicion is a cognitive process based in part on uncertainty – both predisposition to trust and distrust remove some of that uncertainty (P. Bobko et al., 2014) – since, by definition, those states are either certainty of positive or negative outcomes, respectively. Suspicion researchers believe trust may inhibit state-suspicion by deemphasizing Stage I environmental cues, and distrust may act as a catalyst for state-level suspicion (P. Bobko et al., 2014; J. Mayer & Mussweiler, 2011). Since suspicious thought involves the cognitive generation and consideration of multiple plausible, rival hypotheses for the observed behavior (P. Bobko et al., 2014; J. Mayer & Mussweiler, 2011), individual differences were also interesting to consider. For instance, a person who is creative and has "extra" cognitive capacity was believed to be more capable of engaging in suspicious thought while continuing normal operations (P. Bobko et al., 2014). Bobko et al.

proposed that a person's trait-level attributes of creativity, need for cognition, cognitive capacity, and propensity to trust create the trait-level factor "capacity to become suspicious" and serve as antecedents to state-suspicion. Although my research did not study suspicion directly, the test model discussed in Section 3.2.1 supported the collection and analysis of trait-level data through pre-test questionnaires to assess the relationships between individual traits, operator suspicion, and HMT performance.

Stage III of the state-suspicion model referred to potential outcomes (physical manifestations) of suspicion. Increased cognitive activity, as measured by the NASA TLX questionnaire (NASA, 2016), was the physical outcome of interest to my research. Other outcomes were important, but increased cognitive activity has known metrics making it a better outcome measurement. Researchers within the AFRL suspicion portfolio are working on measures for some of the emotional and physiological outcomes such as fear and anxiety. For example, Professor Leanne Hirschfield of Syracuse University is working on measuring suspicion in the brain with functional near-infrared spectroscopy (fNIRS). The AFRL research group also developed a State-Suspicion Index (SSI) questionnaire, which was used in my research. The SSI questionnaire was used to measure a person's levels of uncertainty, perception of malicious intent, cognitive activation, and state-suspicion about a given scenario at a point in time (Philip Bobko, Barelka, Hirshfield, & Lyons, 2014).

2.4 Propositions from Suspicion Research

The research propositions referred to throughout this text were gleaned from a review of the following literature, which spanned multiple social science domains including psychology, human factors, marketing, management, and communication:

- Suspicion leads to suspended judgment
  - (P. Bobko et al., 2014; Hilton, Fein, & Miller, 1993)
- Trust inhibits suspicion; distrust can act as a catalyst for suspicion
  - (P. Bobko et al., 2014; Buller & Burgoon, 1996; Lee & See, 2004; Mcknight, Choudhury, & Kacmar, 2002)
- Suspicion leads to increased cognitive activity
  - (P. Bobko et al., 2014; J. Mayer & Mussweiler, 2011)
- Trait-level attributes / domain knowledge influences one's capacity to become suspicious
  - (P. Bobko et al., 2014; J. Mayer & Mussweiler, 2011)

My research developed questions and hypotheses in Section 2.5 to link these suspicion theories to observable behaviors in an attempt to replicate and explain operator response to cyber-attacks. Section 3 discusses the experimental design developed to answer the questions and hypotheses.

2.5 Problem Definition and Questions / Hypotheses

2.5.1 Problem Definition

There is considerable current effort to prevent or detect and mitigate cyber-attacks on DoD networks and IT systems. In contrast, cyber-physical systems – such as RPAS – represent an intrinsic vulnerability, or at the minimum, a possibility for adversaries to perform cyber-attacks for counter-control or subversion of military assets. As an example, Iranian cyber capabilities were believed to have brought down the Central Intelligence Agency operated RQ-170 Sentinel drone operating near the Iranian border. The Iranians successfully landed the drone in December 2011 causing grave concern over potential compromise of highly sensitive surveillance capabilities. This incident sparked much research directed towards the physical

(hardware / software) security of unmanned vehicle systems. Although much work is being done to study performance of the physical system (Horowitz & Pierce, 2013; Jones & Horowitz, 2012b) in this context, research addressing the human dimension of cyber-attack response from an operator and operational perspective is sparse and represents an emergent area of research needed to fully address cyber-attacks against cyber-physical systems. The questions and hypotheses in Section 2.5.2 start to address this human dimension, and the framework for data collection and analysis was presented in Section 3 with analysis results and finds discussed in Sections 4 and 5. Previous experiments (Gay et al., 2015; Horowitz & Jones, 2015) highlighted the utility of a *Sentinel*-type cyber-attack detection capability; however, operators did not appear to suspect malicious intent and were unsure of their response to *Sentinel* alerts. Therefore, performance against cyber-attacks must be viewed from systems-oriented perspective (i.e. an operator and *Sentinel* team; a.k.a. HMT) with emphasis on the operator's ability to accurately assess and respond to a given situation. My research effort addressed the issue of operator response to cyber-attacks when the operator and *Sentinel* were paired together in an HMT by applying the suspicion theory to scenario based, human-in-the loop, behavioral science experiments involving operators of a representative cyber-physical system in various combinations of cyber / non-cyber contested environments and *Sentinel* alerts received / not received.

2.5.2 Questions and Hypotheses

My research addressed two categories of questions and hypotheses: 1) those related to the application of suspicion theory to operator detection and response to cyber-attacks on unmanned systems and 2) those related to the theory of suspicion itself. The questions and hypotheses related to the application of suspicion theory to operator detection and response to

cyber-attacks on unmanned systems were the primary interest; however, the experimental design offered a unique opportunity to collect and analyze data related to the theory of suspicion itself to inform the suspicion community.

2.5.2.1 Questions

1) The following two questions were related to the application of suspicion theory to operator detection and response to cyber-attacks on unmanned systems and formed the primary focus of this research. I denoted these questions as Focus Questions (FQ).

   1) <u>Focus Question 1 (FQ-1):</u> How does suspicion effect human-machine team (HMT) performance?

   For this study a human-machine team was defined as the pairing of the operator of a cyber-physical system (i.e. an unmanned ground vehicle, UGV) with a *Sentinel* cyber-attack detection aid. The performance consisted of two components, "Score" and "Time," and each was recorded independently for each mission scenario based on the operator's response to that mission. The "Score" reflected the decision-making component of the performance, and "Time" reflected the length of time required to arrive at the decision (operator response time).

   2) <u>Focus Question 2 (FQ-2):</u> How does consequence effect the relationship between suspicion and HMT performance?

   For this study consequence was a two-level factor rated as either Low or High. The factor "consequence" was manipulated through the context of the mission scenario in order to create the Low or High perception of consequence within the operator. For instance, one Low consequence mission scenario was a training mission in the United

States; whereas, one High consequence mission scenario was an operational mission in an undisclosed Middle-Eastern country. The operator's perception of the consequence was measured via post-mission scenario questionnaires.

2) The following questions were related directly to the theory of suspicion and associated propositions as proposed by (P. Bobko et al., 2014). Although secondary to my main research focus, these questions were important to the suspicion community, and my experimental designed allowed for the collection and analysis of data to provide insightful responses to the community. I denoted these questions as Response Questions (RQ).

   1) **Response Question 1 (RQ-1):** What is the relationship between general trait-level attributes and operator suspicion?

      Many traits potentially effect formation of suspicion; however, Bobko et al. discussed creativity, cognitive ability, need for cognition, and propensity to trust as key factors believed to be related to one's "capacity to become suspicious." The experimental design allowed for the collection and analysis of data to provide novel insights concerning the propositions. The trait-level data was collected from each operator using the pre-test questionnaires found in **Appendix II**.

   2) **Response Question 2:** How does perception of consequence affect operator suspicion?

      Trusting in the old adage, "Perception is reality," the experimental design supported collection of data via post-mission scenario questionnaires **(Appendix III)** regarding the operator's perception of the scenario-based (actual) mission consequence. This data was assessed to determine the potential relationship between the scenario-based (actual) consequence, the operator's perception of that consequence, and the operator's suspicion and performance.

2.5.2.2 Hypotheses

1) The following hypotheses were related to the Focus Questions regarding the application of suspicion theory to operator detection and response to cyber-attacks. I denoted these as Focus Hypotheses (FH) and included a set of focus hypotheses for each Focus Question.
   - **FH.1.1:** *Sentinel* alert is related to Operator suspicion.
   - **FH.1.2:** Operator suspicion is positively related to HMT performance.
   - **FH.1.3:** Cyber-attack / *Sentinel* alert combinations are related to operator suspicion.
     - **FH.1.3.a:** No cyber-attack / no *Sentinel* alert
     - **FH.1.3.b:** Cyber-attack / *Sentinel* alert
     - **FH.1.3.c:** No cyber-attack / *Sentinel* alert (False +)
     - **FH.1.3.d:** Cyber-attack / no *Sentinel* alert (False -)
   - **FH.1.4:** Operator suspicion is positively related to operator response time.
   - **FH.2.1:** Consequence alters the direction or strength of the relationship between operator suspicion and HMT performance.
   - **FH.2.2:** Consequence alters the direction or strength of the relationship between operator suspicion and task response time.
2) The following hypotheses were related to the Response Questions regarding the theory of suspicion and associated propositions as proposed by Bobko et al. I denoted these as Response Hypotheses (RH) and included a set of response hypotheses for each Response Question.
   - **RH.1.1:** Creativity is positively related to operator suspicion.
   - **RH.1.2:** Cognitive capacity is positively related to operator suspicion.
   - **RH.1.3:** Propensity to trust is negatively related to operator suspicion.

- **RH.1.4:** Need for cognition is positively related to operator suspicion.
- **RH.2.1:** Operator suspicion mediates (explains) the relationship between perception of consequence and operator performance.
- **RH.2.2:** Operator suspicion mediates (explains) the relationship between perception of consequence and task response time.

Chapter 3: Methodology and Design of Experiment

3.1 Chapter Overview

This chapter introduces the methodology and experimental design implemented to address the research questions and hypotheses from Section 2.5.2. It provides a model depicting the variables for analysis and links the theory of suspicion to the application (i.e. operator detection and response to cyber-attacks on unmanned systems). The chapter also provides a discussion of the test design used to manipulate the factors of interest and the measurement constructs used for data collection.

3.2 Methodology

The primary focus of my research evaluated the relationship between operator suspicion and the detection and response to cyber-attacks on unmanned systems and acknowledged that suspicion was a derivative variable in the test design representing the hypothetical construct of suspicion. Theoretically, suspicion consists of three components: uncertainty, increased cognitive activity, and perception of malicious intent and all three components must occur simultaneously for suspicion to occur (P. Bobko et al., 2014). Therefore, the occurrence of suspicion and its effects must be derived from performance outcome measures. The experimental methodology provided the framework for the collection of multiple pre and post-

test data points to make meaningful observations regarding the relationship between operator suspicion and detection and response to cyber-attacks on unmanned systems, and the design of experiment discussed in Section 3.3 operationalized it through scenario based, human-in-the loop behavioral science experiments with Air Force personnel.

3.2.1 Test Model: Description

Figure 3 provides a depiction of the experiment test model, and it is followed by a brief description of the model elements. As indicated in Section 2.5.2: Questions and Hypotheses, my research addressed two categories of questions and hypotheses: 1) Focus - those related to the application of suspicion theory to operator detection and response to cyber-attacks on unmanned systems and 2) Response - those related to the theory of suspicion itself. The model was designed to address both the Focus and the Response areas. The factors contributing to the analysis of the Response hypotheses were in the gray shaded part of the model on the left-hand side of the orange Suspicion oval. The factors contributing to the Focus hypotheses were in the blue shaded part of the model on the right-hand side of the orange Suspicion oval.

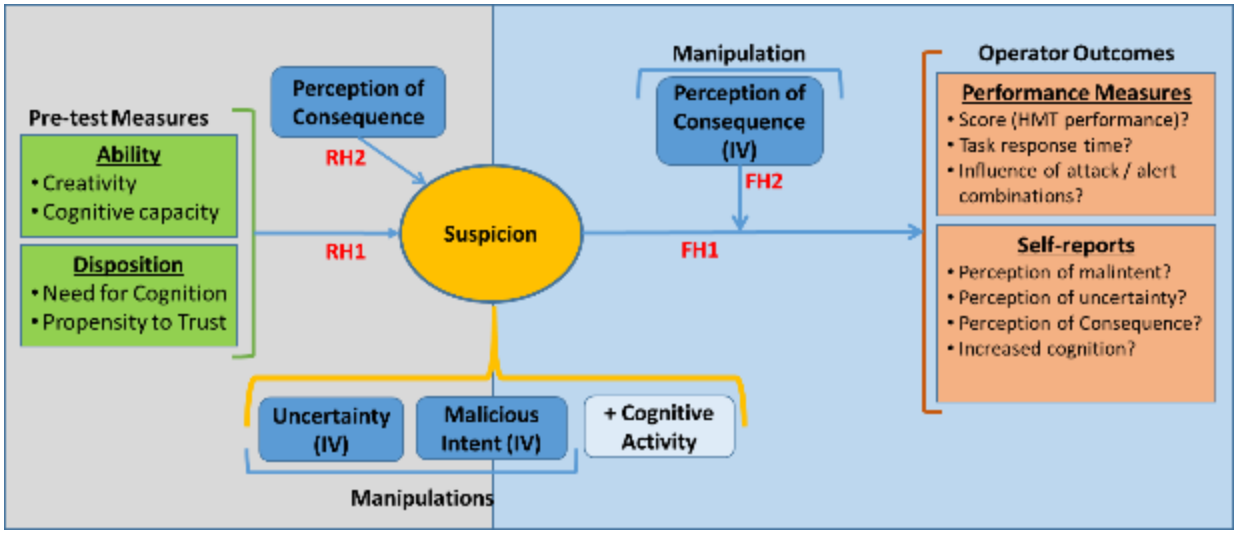


Figure 3: Test Model for Operator Suspicion Experiment

A description of the model elements is provided below.

- Orange oval – Suspicion: The orange oval at the center of the model was considered a derived variable representing the hypothetical construct of Suspicion; therefore, Suspicion (and its effect) must be evaluated from the Operator Outcomes listed in the salmon colored box on the right-hand side of the model.
- Blue boxes – Independent Variables (IV): The model was based on a three factor, two level design. The model contained three IV's represented by the dark blue boxes, and each IV had the two levels: Low and High. The IV's of Uncertainty and Malicious Intent represent two of the three components of the suspicion theory. The third IV in the model was Consequence. Although Consequence was not a component of the suspicion theory, it was believed to affect the relationship between operator Suspicion (orange oval in center or model) and Operator Outcomes (salmon colored box on the right-hand side of the model). All three IV's were manipulated Low or High through the context of the mission scenarios (**Appendix I**). Consequence occurred at two places on the model representing two different analytical relationships – moderation and meditation – which were discussed in Section 3.2.2.1 and 3.2.2.2.
- Light Blue box – + Cognitive Activity: The third component of suspicion was +Cognitive Activity. This variable represented the operator's increased cognitive load due to interaction with the mission scenario, and it was measured at the end of each scenario using the NASA TXL and State-Suspicion Index (SSI) questionnaires found in **Appendix III**.
- Salmon colored box – Operator Outcomes: Operator Outcomes were measured in two ways: Performance Measures and Self-reports. "Score" and "Time" were the primary

variables for operator performance measures. Time data was recorded within the TurningPoint software package (Turning Techologies, 2013) used to interface with the experiment, and Score data was determined post-experiment by evaluating the operator's decision tree sequence (e.g. **Figure 7**, Section 3.3.3) logged in the TurningPoint software against a scoring rubric developed with subject matter expert input. Each scenario had its own unique Score rubric (e.g. **Table 6**, Section 3.3.4). "Score" and "Time" variables were reflective of the operator's performance against the actual sequence of events in the experiment. Self-report outcomes were collected via questionnaires at the end of each mission scenario **(Appendix III)**. Collectively, these questionnaires assessed the operator's perception of Uncertainty, Malicious Intent, Cognitive Activity, and Consequence as a result of the mission scenario just completed.

- Green box – Pre-test Measures: The green box on the left-hand side of the test model represented pre-test measures assessed for each test subject prior to their start of the experiment. Bobko et al. proposed Creativity, Cognitive Capacity, Need for Cognition, and Propensity to Trust as four attributes potentially linked to one's "capacity to become suspicious." Although the relationship of these factors to operator suspicion was not the primary focus of my research, the test model supported data collection and analysis of these attributes through the use of pre-test questionnaires (**Appendix II**). Correlation of pre-test attributes with post-test operator outcomes was accomplished. This provided a unique opportunity to provide a response to the suspicion community regarding Bobko et al. trait-level propositions.

3.2.2 Test Model: Analysis Approach

Each arrow in the test model shown in **Figure 3** represented a method of analysis to address the associated Focus or Response questions and hypotheses, which were designated in the model as FH and RH, respectively. The test subjects (operators) in the experiment were Air Force officers at the Air Force Institute of Technology (AFIT). Each operator was exposed to the same set of eight different mission scenarios over a two-hour period and data was collected on the operators' responses to each of the mission scenarios. This data, which was repeatedly gathered on the operators, was hierarchical in nature, as all the observations were nested within the individuals (Osborne, 2000; Woltman, Feldstain, MacKay, & Rocchi, 2012). Nesting in hierarchical data creates an issue for analysis in that the normal assumptions of independence required by most analytical methods are violated due to the shared characteristics of the individuals, and the resulting ordinary least squares regression produces standard errors that are too small (Osborne, 2000), which may erroneously lead one to believe an effect or relationship exists. Since the data was hierarchical in nature, the preferred method of analysis to overcome this lack of independence (shared variance) was hierarchical linear modeling (HLM). HLM is a complex form of ordinary least squares regression that is used to analyze the shared variance in the outcome variables when the predictor variables are at varying hierarchical levels thus making it more efficient at accounting for variance among variables at different levels than other existing analyses methods. (Woltman et al., 2012). The data in my experiment was represented by two hierarchical levels: Level-1, Scenario level and Level-2, Operator level. **Table 1** contains examples of factors at each hierarchical level of the experiment.

Hierarchical Level	Example of Hierarchical Level	Example Variables
Level - 2	Operator Level	Creativity
		Cognitive Capacity (GPA)
		Need for Cognition
		Propensity to Trust
		Age
		Gender
Level - 1	Scenario Level	Perception of Uncertainty
		Perception of Malicious Intent
		Perception of Consequence
		Increased Cognition
		Score*
		Time*
		Suspicion*
*The outcome variable is always a Level - 1 variable.		

Table 1: Factors at each hierarchical level

3.2.2.1 Analysis Approach to Focus Hypotheses

The following hypotheses were related to the Focus Questions regarding the application of suspicion theory to operator detection and response to cyber-attacks unmanned systems. The Focus Hypotheses were denoted as FH1 & FH2 in the model (**Figure 3**), and they represented the set of Focus Hypotheses for each Focus Question. A top-level description of the analysis process was provided for each Focus Hypotheses. The results of the analysis were discussed in Chapter 4.

- **FH.1.1:** *Sentinel* alert is related to operator suspicion.

  A new variable "SENTINEL" was created which coded all of the eight scenarios containing a *Sentinel* alert with the value of 1 and those scenarios without a *Sentinel* alert with the value 0. Total Suspicion was denoted as the variable "SSI_TOTAL". In the HLM analysis for this hypothesis, "SENTINEL" was the predictor variable and

"SSI_TOTAL" was the outcome variable. Both variables were centered on the group mean (group centered) and occurred at Level-1 of the hierarchy.

- **FH.1.2:** Operator suspicion is positively related to HMT performance.

  In the HLM analysis of this hypothesis "SSI_TOTAL" was the predictor variable and "Score" was the outcome variable. Both variables were group centered and occurred at Level-1 of the hierarchy.

- **FH.1.3:** Cyber-attack / *Sentinel* alert combination is related to operator suspicion.

  Four Cyber-attack / *Sentinel* alert combinations existed, and two of each Cyber-attack / *Sentinel* alert combinations were represented in the set of eight mission scenarios. Thus, the first step in this HLM analysis process was the creation of four new group centered Level-1 variables to represent these combinations. The new variable names **"BOLD"** were included in the following sub-hypotheses description.

  - **FH.1.3.a:** No cyber-attack / no *Sentinel* alert, **"NA_NA"**

    In the HLM analysis of this hypothesis "NA_NA" was the predictor variable and "SSI_TOTAL" was the outcome variable.

  - **FH.1.3.b:** Cyber-attack / *Sentinel* alert, **"SE_COR"**

    In the HLM analysis of this hypothesis "SE_COR" was the predictor variable and "SSI_TOTAL" was the outcome variable.

  - **FH.1.3.c:** No cyber-attack / *Sentinel* alert (False positive, F +), **"SE_T1E"**

    In the HLM analysis of this hypothesis "SE_T1E" was the predictor variable and "SSI_TOTAL" was the outcome variable.

- **FH.1.3.d:** Cyber-attack / no *Sentinel* alert (False negative, F -), **"SE_T2E"**

  In the HLM analysis of this hypothesis "SE_T2E" was the predictor variable and "SSI_TOTAL" was the outcome variable.

Although not directly related to FH.1.3, I conducted additional HLM analysis of the outcome variables "Score" and "Time" with each of these combinations. I also considered the interaction of these combinations with "SSI_TOTAL" as predictors of "Score" and "Time."

- **FH.1.4:** Operator suspicion is positively related to operator response time.

  In the HLM analysis of this hypothesis "SSI_TOTAL" was the predictor variable and "Time" was the outcome variable. Both variables were group centered and occurred at Level-1 of the hierarchy.

- **FH.2.1:** Consequence alters the direction or strength of the relationship between operator suspicion and HMT performance.

  The purpose of this analysis was to determine if the variable for consequence ("CON1") was a moderator of the relationship between operator suspicion ("SSI_TOTAL") and operator performance ("Score"). A moderator is a variable that alters the direction or strength of the relationship between a predictor and an outcome (Baron & Kenny, 1986); thus, it is an interaction whereby the effect of one variable depends on the level of another (Frazier, Tix, & Barron, 2004). I assessed the moderator effect by using three variables and three lines of analysis. I used the outcome variable "Score," the predictor variable "SSI_TOTAL," and the moderator variable "CON1". The three lines of analysis were: predictor to outcome, moderator to outcome, and the product of the predictor and moderator (predictor x moderator) to outcome (Baron & Kenny, 1986; Frazier et al.,

2004). The first step in the HLM analysis was the creation of a moderation product variable ("CONSMOD"), which was "SSI_TOTAL" x "CON1". All of these variables were group centered and occurred at Level-1 of the hierarchy. HLM analysis was run using "SSI_TOTAL," "CON1," and "CONSMOD" as predictor variable inputs to determine the outcome variable "Score". Consequence was supported as a moderator variable if the "CONSMOD" interaction was significant (Baron & Kenny, 1986).

- **FH.2.2:** Consequence alters the direction or strength of the relationship between operator suspicion and task response time.

  The purpose of this analysis is to determine if the variable for consequence ("CON1") was a moderator of the relationship between operator suspicion ("SSI_TOTAL") and task response time ("Time"). The same analytical procedures discussed in FH.2.1 (above) were applied; however, "Time" was used as the outcome variable in the analysis.

3.2.2.2 Analysis Approach to Response Hypotheses

The following hypotheses were related to the response questions regarding the theory of suspicion and associated propositions as proposed by Bobko et al. The Response Hypotheses were denoted as RH1 & RH2 in the model (**Figure 3**), and they represented the set of Response Hypotheses for each Response Question. A top-level description of the analysis process was provided for each Response Hypotheses. The results of the analysis were discussed in Chapter 4.

- **RH.1.1:** Creativity is positively related to operator suspicion.

  In the HLM analysis of this hypothesis operator suspicion ("SSI_TOTAL") was the outcome variable, and it was a Level-1 group centered variable. Creativity

("CREATIVITY") was the predictor variable, and it was a Level-2 variable center on the grand mean (grand centered).

- **RH.1.2:** Cognitive capacity is positively related to operator suspicion.

  In the HLM analysis of this hypothesis operator suspicion ("SSI_TOTAL") was the outcome variable, and it was a Level-1 group centered variable. Cognitive capacity ("GPA_U" and/or "IQ1") was the predictor variable, and it was a Level-2 grand centered variable.

- **RH.1.3:** Propensity to trust is negatively related to operator suspicion.

  In the HLM analysis of this hypothesis operator suspicion ("SSI_TOTAL") was the outcome variable, and it was a Level-1 group centered variable. Propensity to trust ("TRUST_MA") was the predictor variable, and it was a Level-2 grand centered variable.

- **RH.1.4:** Need for cognition is positively related to operator suspicion.

  In the HLM analysis of this hypothesis operator suspicion ("SSI_TOTAL") was the outcome variable, and it was a Level-1 group centered variable. Need for cognition ("NCOG") was the predictor variable, and it was a Level-2 grand centered variable.

- **RH.2.1:** Operator suspicion mediates (explains) the relationship between perception of consequence and operator performance.

  The purpose of this analysis was to determine if operator suspicion ("SSI_TOTAL") was a mediator of the relationship between perception of consequence ("CON1") and operator performance ("Score"). Mediators establish "how" and "why" one variable predicts or influences an outcome variable by explaining the relationship and mechanism through which a predictor influences an outcome variable (Baron & Kenny, 1986; Frazier et al., 2004). I assessed the mediator effect by using a method developed by Kenny and his

colleagues. This method was thought to be the most common method for testing mediation in psychological research, and it required four steps (performed with three regression equations) to establish that the variable suspicion ("SSI-TOTAL") mediates the relationship between the predictor variable perception of consequence ("CON1") and the outcome variable operator performance ("Score") (Baron & Kenny, 1986; Frazier et al., 2004). I've provided a graphical representation of that process in **Figure 4** below.

- **RH.2.2:** Operator suspicion mediates (explains) the relationship between perception of consequence and task response time.

  The purpose of this analysis was to determine if operator suspicion ("SSI_TOTAL") was a mediator of the relationship between perception of consequence ("CON1") and task response time ("Time"). The same analytical procedures discussed in RH.2.1 (above) were applied; however, "TIME" was used as the outcome variable in the analysis.

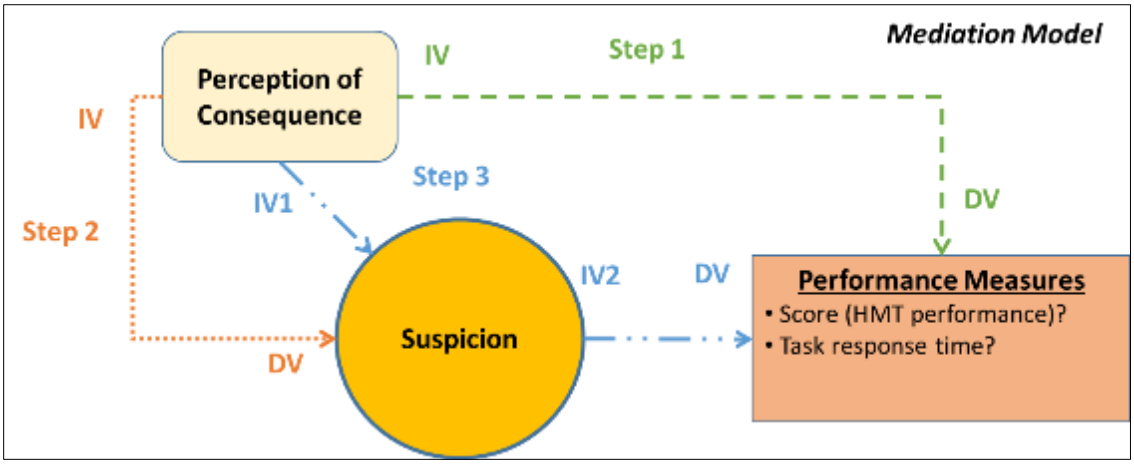


Figure 4: Analysis Model for Mediation

3.3 Design of Experiment (DoE)

The primary focus of my research was to evaluate the relationship between operator suspicion and the detection and response to cyber-attacks on unmanned systems while

acknowledging that suspicion is a derivative variable in the test design representing the hypothetical construct of suspicion. Theoretically, suspicion consists of three components: uncertainty, increased cognitive activity, and a perception of malicious intent. All three components must occur simultaneously for suspicion to occur (P. Bobko et al., 2014). Therefore, the occurrence of suspicion and its effects must be derived from performance outcome measures. The experimental methodology presented in Section 3.2 provided the framework for the collection of multiple pre and post-test data points to make meaningful observations regarding the relationship between operator suspicion and detection and response to cyber-attacks on unmanned systems, and the design of experiment (DoE) operationalized it through scenario based, human-in-the loop, behavioral science experiments with Air Force personnel. In order to collect statistically meaningful data, the DoE accomplished three main task: 1) it effectively addressed many known threats to experimental validity, 2) it constructed test scenarios that accurately reflect the theory of suspicion in the context of interest, and 3) it provided a realistic method of operationalizing the test scenarios to allow for data collection and analysis. In order to address these tasks, a 3-factor, 2-level, with-in subjects, repeated measures DoE was implemented. This DoE was discussed in the following sections.

3.3.1 DoE: Threats to Validity

When designing the experiment it was critical to ensure the validity of the inferences about the higher order constructs of interest (Shadish, Cook, & Campbell, 2002) and to show that the experiment produced results that were consistent with the construct (Sackett & Larson, 1990). In other words, did the experiment measure the intended construct? Suspicion theory, as proposed by Bobko et al., was the construct of interest in this experiment, and the experiment was specifically designed to manipulate and measure the elements of that theory. The three

elements of the suspicion theory are uncertainty, a perception of malicious intent, and cognitive activation. The elements of "uncertainty" and "perception of malicious intent" were treated as two of the factors in the DoE and directly manipulated Low or High through the scenarios. Although not an element of suspicion, the potential effect of "perception of consequence" on operator decision-making was an interest item in the study. Therefore, "consequence" became the third factor in the DoE, and it was also manipulated Low or High through the scenarios. Cognitive activation – the third element of suspicion – was not directly manipulated, but its affect was measured. Bobko et al. developed the original twenty-item state suspicion index (SSI) to "generally" measure these elements of suspicion, and I worked directly with Dr. Bobko to tailor it into a contextually relevant thirteen-item SSI measure for my research. The new thirteen-item SSI measure received a Cronbach's alpha (reliability rating) of 0.881. Additionally, separate manipulation checks and pilot study experiments were conducted prior to the start of the main experiment to ensure the construct of suspicion was, in fact, being measured. Other constructs (e.g. propensity to trust, need for cognition, and creativity) were also measured pre-test in order to support my response hypotheses associated with Bobko et al. propositions regarding individual predisposition to become suspicious. The questionnaire measurements used in this study were derived from published literature, and they were listed in **Table 2** below. The questionnaires were include in **Appendix II & III**.

Measurement Questionnaire	Construct	Source	Number of Items	Scale	Cronbach's Alpha (Reliability)
Trust	Propensity to trust	(R. C. Mayer, Davis, & Schoorman, 1995)	8	Likert (1-7)	0.752
Need for Cognition	General need for cognition	(J. T. . Cacioppo, Petty, & Kao, 1984) (J. T. Cacioppo, Petty, Feinstein, & Jarvis, 1996)	18	Likert (1-7)	0.866
Creativity	General creativity	Bobko's suspicion research	2	Likert (1-4)	0.570
State Suspicion Index (SSI)	Perception of suspicion, uncertainty, malicious intent, and cognitive activation	Co-developed with Bobko from original SSI (Philip Bobko et al., 2014)	13	Likert (1-7)	0.881
NASA Task Load Index (TLX)	Factors related to cognitive workload	NASA TLX website	6	0-100	0.839

Table 2: Measurement Questionnaires and Reliabilities

The DoE also considered threats to internal validity. Internal validity is concerned with inferences about the causal relationships between the independent and dependent variables (Sackett & Larson, 1990; Shadish et al., 2002). In other words, is the observed effect in the study due to the manipulation of the independent variables or some other factors of the experiment? Carryover effects are common threats to internal validity of with-in subject, repeated measure designs. The principle of carryover effect recognizes the fact that exposure to one manipulation or test could have persistent consequence that affect the result of subsequence tests in a with-in subject, repeated measure design, and common types of carryover effects include order effect, practice / learning effect, and fatigue effect (Neale & Leibert, 1986). Since this study was a with-in subject, repeated measures design where each test participant was exposed to a series of eight different mission scenarios, I had to safe guard against it. Randomization and counterbalancing are the two most widely used methods for countering carryover effects (Kirk, 1995; Neale & Leibert, 1986; Shadish et al., 2002), and I employed both

of those techniques in my DoE. As shown in **Table 3** below, the experimental design consisted of two test cases (Test Case #1 and Test Case #2), and each test case consisted of eight mission scenarios listed as "Standard Order" (1-8).

Three Factor x Two Level Test Sequence (Sentinel) - as of 6-29-16				Test Case #1	
Standard Order	Factor A: Uncertainty	Factor B: Perception of Malintent	Factor C: Consequence	Attack	Sentinel Response Received
1	Lo (7% prob of system issue)	Lo (CONUS)	Lo	N	Y
	Resupply Mission, Pre-deployment Qualification				
2	Hi (50% prob of system issue)	Lo (CONUS)	Lo	Y	Y
	Remote Resupply / Trans Delivery Mission, Flag Exercise				
3	Lo (7% prob of system issue)	High (AOR)	Lo	N	Y
	Routine Operational Resupply				
4	Hi (50% prob of system issue)	High (AOR)	Lo	Y	Y
	Remote Operational Resupply Mission				
5	Lo (7% prob of system issue)	Lo (CONUS)	Hi	N	N
	Transport of Nuclear Material, Joint Exercise with DOE				
6	Hi (50% prob of system issue)	Lo (CONUS)	Hi	Y	N
	Transport of Nuclear Material, Operational Joint Mission with DOE				
7	Lo (7% prob of system issue)	High (AOR)	Hi	N	N
	Operational Resupply of Supported Unit				
8	Hi (50% prob of system issue)	High (AOR)	Hi	Y	N
	Remote Resupply / Transportation Delivery of SOF Unit				

Three Factor x Two Level Test Sequence (Sentinel) - as of 6-29-16				Test Case #2	
Standard Order	Factor A: Uncertainty	Factor B: Perception of Malintent	Factor C: Consequence	Attack	Sentinel Response Received
1	Lo (7% prob of system issue)	Lo (CONUS)	Lo	Y	N
	Resupply Mission, Pre-deployment Qualification				
2	Hi (50% prob of system issue)	Lo (CONUS)	Lo	N	N
	Remote Resupply / Trans Delivery Mission, Flag Exercise				
3	Lo (7% prob of system issue)	High (AOR)	Lo	Y	N
	Routine Operational Resupply				
4	Hi (50% prob of system issue)	High (AOR)	Lo	N	N
	Remote Operational Resupply Mission				
5	Lo (7% prob of system issue)	Lo (CONUS)	Hi	Y	Y
	Transport of Nuclear Material, Joint Exercise with DOE				
6	Hi (50% prob of system issue)	Lo (CONUS)	Hi	N	Y
	Transport of Nuclear Material, Operational Joint Mission with DOE				
7	Lo (7% prob of system issue)	High (AOR)	Hi	Y	Y
	Operational Resupply of Supported Unit				
8	Hi (50% prob of system issue)	High (AOR)	Hi	N	Y
	Remote Resupply / Transportation Delivery of SOF Unit				

Table 3: Experimental Design with Counterbalancing

Table 3 also indicates the eight mission scenarios (1-8) were different within a test case in order to achieve the desired manipulations; however, the scenarios (1-8) were the same across each test case. In other words scenario "Standard Order 1" in "Test Case 1" was the same as scenario "Standard Order 1" in "Test Case 2." Randomization occurred both within each test case and

between each test case. Test subjects were randomly assigned (alternating) an order in which they would encounter the mission scenarios as they arrived to participate in the experiment. The assigned orders were either "Standard Order" sequence 4 to 3 (i.e. 4,5,6,7,8,1,2,3) or "Standard Order" sequence 3 to 4 (i.e. 3,2,1,8,7,6,5,4). Additionally, test subjects were randomly assigned to either Test Case #1 or Test Case #2 (alternating) as they arrived to participate in the experiment. Counterbalancing occurred between the two test cases to balance the effect of "Cyber-Attack" (yes / no) and "Sentinel Response Received" (yes / no). This was shown in the last two columns of "Test Case #1" and "Test Case #2" of **Table 3**. This DoE controlled for potential carryover effects through the implementation of these randomization and counterbalancing techniques.

Generalizability was another potential threat addressed by the DoE. It is often viewed as an extension of external validity, and it is concerned with inferences about the extendibility of the causal relationships to other times, settings or individuals (Sackett & Larson, 1990; Shadish et al., 2002). According to Sackett, generalizability is a function of methodology, not results, and the degree to which outcomes can be generalized is either built into or out of the experimental design. Methodological choices pertaining to the participants in the study, the setting in which the research is conducted, and operationalization of the variables of interest are key decisions impacting generalizability (Sackett & Larson, 1990). The motivation for this study stemmed from observations gained while conducting simulation experiments with Air Force personnel involved in the operation of remotely piloted aircraft systems (RPAS) in a cyber contested environment. Two factors made this environment near impossible to replicate operationally for my study: 1) RPAS operators and systems have an extremely high operational demand and 2) cyber-attacks are unpredictable. Given these two factors and the need to establish

a certain amount of experimental control, I made the next best choice according to Sackett and selected a representative sample of participants and experimental platform. I chose Air Force personnel at the Air Force Institute of Technology as my test subjects, and I used an unmanned ground vehicle (UGV) system as a surrogate test platform. I also referred to (Cohen, 1992) to get an estimate of the sample size recommended for statistical relevance based on a power of 0.80, a medium to large effect size, and an α of 0.05. Based on **Table 4** from Cohen's work, I endeavored to recruit 34 to 76 Air Force personnel for the experiment.

N for Small, Medium, and Large ES at Power = .80 for α = .01, .05, and .10

	α								
	.01			.05			.10		
Test	Sm	Med	Lg	Sm	Med	Lg	Sm	Med	Lg
8. Mult *R*									
2k[b]	698	97	45	481	67	30			
3k[b]	780	108	50	547	76	34			
4k[b]	841	118	55	599	84	38			
5k[b]	901	126	59	645	91	42			
6k[b]	953	134	63	686	97	45			
7k[b]	998	141	66	726	102	48			
8k[b]	1,039	147	69	757	107	50			

Note. ES = population effect size, Sm = small, Med = medium, Lg = large, diff = difference, ANOVA = analysis of variance. Tests numbered as in Table 1.
[a] Number of groups. [b] Number of independent variables.

Table 4: Statistical Power (Jacob Cohen)

These choices allowed me to study the construct of suspicion in a controlled environment using a representative sample and a platform which allowed for control of the exposure to cyber-attacks and system (*Sentinel*) alerts. Additionally, the mission scenarios contained in **Appendix I** were constructed in the format of an operational mission brief, and the mission content was representative of real mission sets. Finally, many operations associated with RPAS missions occur in an office type environment using standard office equipment such as computers and monitors. The experimental setting was an office environment with computer and monitor

equipment. The experimental task was to monitor the mission track, video and instrument readouts from the UGV mission and respond to system anomalies (e.g. cyber-attack or other), and these tasks were closely aligned with traditional RPAS tasks. Many efforts were made to safeguard against issues concerning generalizability, but "…*generalizability cannot be guaranteed. Because future events can never be represented in current samples, generalizability across time is always a matter of faith*" (Sackett & Larson, 1990).

3.3.2 DoE: Scenario Development

Mission scenarios were carefully constructed to accurately reflect the theory of suspicion in a relevant military context. The mission scenarios consisted of two components: 1) the mission briefing and 2) the mission video (discussed in Section 3.3.3: DoE Operationalization). This section focused on development of the mission briefings. The mission briefings contained in **Appendix I** were created to resemble standard military mission briefings in both format and content. They each contained three main bulleted sections and four sub-bulleted topics. The three main bulleted sections were Mission ID, Mission Location, and Mission Briefing, and the Mission Briefing section contained the four sub-bulleted topics of Description, Threat Environment, Likelihood of Mission Success, and Risk. The content of each section worked in concert to create the desired manipulation effect indicated in **Table 3**.

Mission location was one variable used in the manipulation of perception of malicious intent, and the Mission Location section consisted of two possible locations: the United Stated or an Undisclosed Middle Eastern Country. Missions set in the United States were intended to engender a lower perception of malicious intent; whereas, those occurring in a Middle Eastern Country were intended to provoke a higher perception of malicious intent. Additional information concerning the type of mission, adversary actions in the objective area, likelihood of

mission success, and potential risks was provided to the operator through the context of the scenario to aid in achieving the desired manipulation effect.

The probability of mission success was one variable used to manipulate the operator's perception of uncertainty about the successful completion of the mission, and the Likelihood of Mission Success section contained one of two probabilistic outcomes. The outcome, one out of fifteen (7%) was indicative of low uncertainty; whereas, ten out of twenty (50%) was suggestive of high uncertainty. It is important to note the use of "uncertainty" in this experiment was strictly for the manipulation and influence of the operator's subjective perception of uncertainty, and it was not intended as a validated statistical measure of uncertainty. The scope of this experiment did not include statistical validation of operator uncertainty. Additional information concerning the location and type of mission, adversary actions in the objective area, and potential risks was provided to the operator through the context of the scenario to help achieve the desired manipulation effect.

Mission type was one variable used to influence the operator's perception of the consequences related to his / her decisions during the mission. All of the missions were categorized as either ground transport or resupply missions; however, the characteristics of the mission and cargo gave some indications as to the consequence of operator actions during the mission. For instances, missions characterized as routine or training with standard cargo such as food and water were intended to engender a feeling of low decision consequence; whereas, missions characterized as operational with hazardous cargo or specialized mission equipment were designed to create a perception of high decision consequence. Additional information concerning the mission location, adversary actions in the objective area, threats and potential

risks was provided to the operator through the context of the scenario to help achieve the desired manipulation effect.

3.3.3 DoE: Operationalization

The test methodology and DoE were of no effect unless they were operationalized in a meaningful way. This may seem like a trivial task, but I assure you it was not. The mission briefings mentioned in Section 3.3.2: DoE – Scenario Development, the mission videos, the test and scoring protocol, and the training materials were all key developments in this task.

The mission videos created the interactive environment for the operator and helped bring the mission briefing context (and manipulations) to life. Each of the two test cases presented in **Table 3** contained eight mission scenarios and eight corresponding mission videos. The videos implemented the "Attack" and "*Sentinel* Response Received" combinations found in each test case. Since counterbalancing was used between the test cases, a total of sixteen mission videos were developed. The mission videos were created using a radio controlled truck as the unmanned ground vehicle (UGV) system – to representative a more generalizable cyber-physical system – and screen capture and video editing software. The UGV system consisted of an UGV, a laptop ground control station (GCS) with Mission Planner (ArduPilot, 2016) mission planning software, a wireless network, modem, and radio controller (RC). The RC was for manual control of the UGV, and it was used by the researcher to initiate an aspect of a cyber-attack (pause vehicle). The UGV was equipped with a GPS, autopilot, RC receiver, modem, power bus, and camera to enable autonomous operations, and it was capable of being controlled by a secondary laptop GCS simultaneously. The second GCS was utilized to initiate a cyber-attack on the speed control parameters of the UGV. CamStudio (CamStudio, 2013) and TinyTake (MangoApps, 2016) screen capture software programs were used to record the missions for playback to the test

subjects (operators), thus reducing test variability, and Filmora Video Editor (Wondershare, 2016) software was used to pixelate the camera image and overlay the *Sentinel* alert response on the screen-captured video. This setup enabled the UGV to run all of the missions in an autonomous mode. It is depicted in **Figure 5** and was used to run and record all of the mission videos.

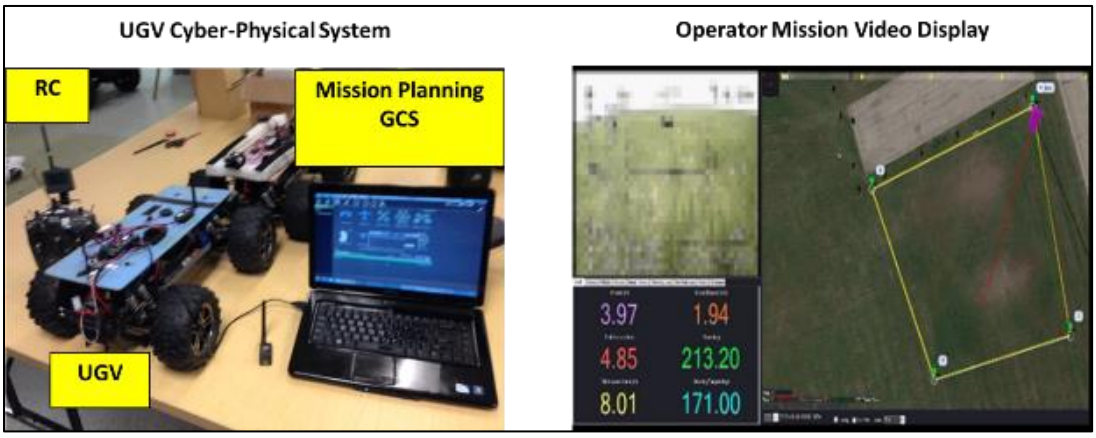


Figure 5: Mission Video Setup

All missions were planned and implemented with the UGV in the autonomous or "auto" mode of operation and the screen capture software was used to record the mission display exactly as the mission unfolded. Since the type of cyber-attack was not intended to be a separate factor in the study, I decided to utilize one attack vector. The attack vector implemented in the study was a "cyber-attack" on the UGV's throttle control, and it was implemented in one of two ways. One throttle control "attack" caused the UGV to pause for 15 seconds, and the other caused the UGV to accelerate beyond its normal speed range. I implemented a simple (easily detectable) type of a "cyber-attack" for this study. I reasoned that if suspicion plays a role in detection and response in this simple type of "cyber-attack", it would likely play a role in a more complex cyber-attack.

In order for the UGV to operate in autonomous mode, a switch on the RC was set to "auto" mode. The pause "cyber-attack" was implemented while the UGV was running the mission in "auto" mode by physically flipping the RC switch from "auto" to "manual" mode which resulted in the UGV stopping and waiting for a manual command from the RC. When the desired time of the "cyber-attack" induced stop passed, the switch on the RC was then toggled back to "auto" mode, and the UGV continued autonomously as it was programmed.

The other throttle control "attack" utilized the secondary GCS to change the UGV speed control parameter settings. The secondary GCS used the mission planning software and the wireless network to write and send new speed control parameters to the UGV during the mission to override the "auto" mode programmed settings. As an example, the standard UGV speed control setting for all missions was 2 meters per second (m/s); however, the secondary GCS sent a command that changed the speed setting to 5 m/s causing the UGV to accelerate beyond the expected speed parameters.

Whenever the mission scenario called for a *Sentinel* response, the phrase, "Cyber Attack: Throttle Control," was overlaid in red text on the lower portion of the pixelated UGV camera image (**Figure 6**). This was accomplished during video editing of the screen-capture mission video. This message remained visible for 30 seconds, and it then faded away. The mission videos implemented the combinations of "Attack" and "*Sentinel* Response Received" found in **Table 3**, and they worked in concert with the mission briefing to create a realistic and engaging environment for testing operator detection and response to "cyber-attacks" on unmanned systems.

Figure 6: Mission Video with Sentinel Alert

The test and scoring protocol was developed using the screen captured mission videos and the mission briefings. The operator's perception of the mission scenario was informed by the mission briefing and the unfolding of events in the mission video. The mission briefing created the framework for understanding the nature of the mission, a priori, and it provided context for the operator's initial assessment of what to expect during the mission. For each of the eight mission scenarios, the test protocol required the operator to read the mission briefing, monitor the "screen captured" mission video, record the UGV speed every 30 seconds, and respond to any abnormal system behaviors that may occur during the mission. Abnormal system behaviors referred to anything the operator observed that appeared different from what was expected. For example, the mission briefing states the UGV should stop at Waypoint 3 for 15 seconds to simulate offloading of supplies; however, the UGV proceeded through Waypoint 3 without stopping and continued on the planned autonomous route. In this case, proceeding through Waypoint 3 was considered abnormal behavior because it deviated from the expected behavior set by the mission briefing.

The "Operator Decision Tree" depicted in **Figure 7** was a flow chart developed for operator training and mission execution to control for variability in operator response options. The decision tree flow chart in **Figure 7** was read from the top down, and the options for the

decision sequence followed the direction of the arrows. Therefore, the operator's first response in the sequence is always "1 – Acknowledge" to indicate his / her belief that something in the mission scenario was different than expected. Based on the decision tree, the following response sequences were possible: 1-1, 1-2-1, 1-2-2, 1-2-3, 1-3-1, and 1-3-2. Some mission scenarios yielded no system anomalies, and thus no "1 – Acknowledge" decision, and these situations resulted in "No Response" from the operator.

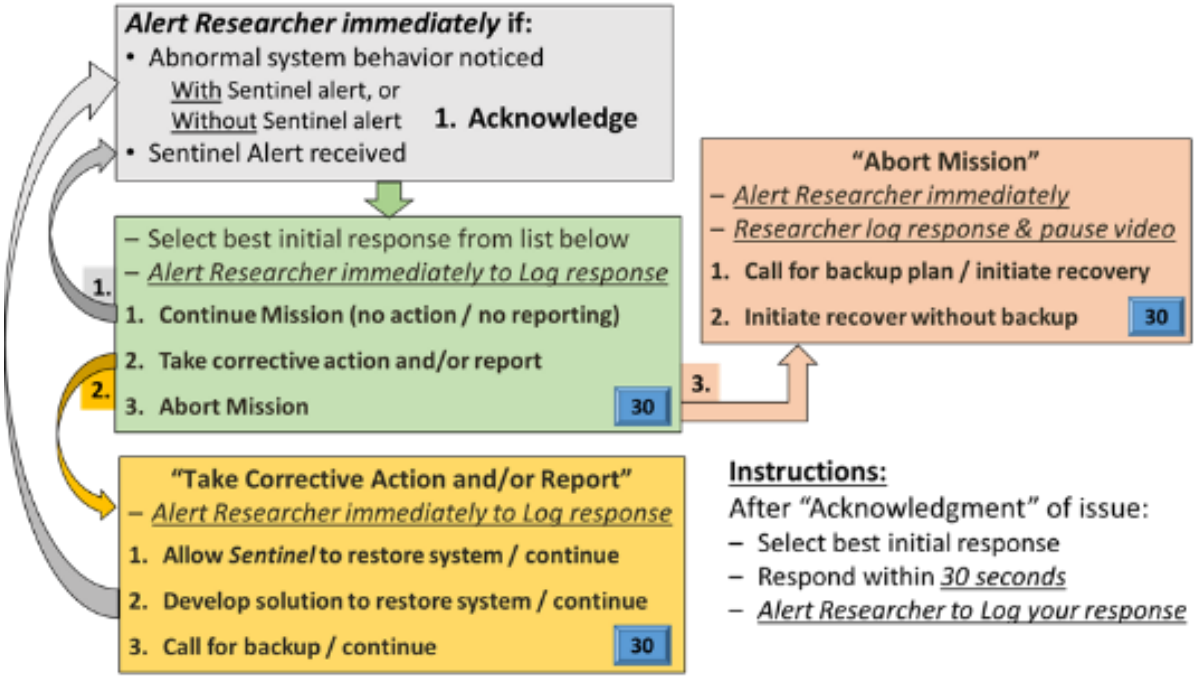


Figure 7: Operator Decision Tree

The "Key to Operator Decision Tree Responses" and the "ResponseCard" shown in **Table 5** were utilized by the researcher to log the operator's response sequences and score operator performance. It was read from left to right and corresponded to the options for the decision sequences presented to the operator via the flow chart in **Figure 7**.

KEY to Operator Decision Tree RESPONSES			
1	Acknowledge		
	1	Continue Mission	
	2	Take corrective action and/or report	
		1	Allow Sentinel to restore / continue
		2	Develop solution to restore / continue
		3	Call for backup / continue
	3	Abort Mission	
		1	Call for backup / initiate recovery
		2	Initiate recover without backup

Table 5: Key to Operator Decision Tree Responses and ResponseCard

Based on the decision tree, the following response sequences were possible: 1-1, 1-2-1, 1-2-2, 1-2-3, 1-3-1, and 1-3-2. Some mission scenarios yielded no system anomalies, and thus no "1 – Acknowledge" decision, and these situations resulted in "No Response" from the operator. The operator was trained to verbalize the response sequence corresponding to his / her decision, and the researcher entered that response sequence into the TurningPoint software via the ResponseCard shown in **Table 5**. The TurningPoint software logged each response in the response sequence and the mission time associated with the response. These data logs were later used to determine response time and performance scores as discussed in Section 3.3.4: DOE: Lexicon and Scoring Approach.

The final stage of operationalization was developing the training materials to allow for consistent implementation of the experiment over multiple experimental runs and operators. The experiment was conducted in three phases: Phase 1- Pre-experiment, Phase 2-Training, and Phase 3-Missions.

Phase 1 – Pre-experiment training materials included a consent form to introduce the test subject to the experiment and gain their concurrence to participate prior to beginning. It then discussed the outline of the experiment and addressed each of the three phases. After the consent

form and outline of the experiment were discussed, the researcher paused to execute the Phase 1-Pre-experiment questions. Phase 1 was a self-paced process in which the operator completed a series of four questionnaires regarding demographic and personality related information **(Appendix II)**. At the conclusion of Phase 1, the researcher entered Phase 2 – Training.

Phase 2 – Training consisted of four blocks: Mission Context, the "M" in HMT, Anatomy of a Mission, and Practice Scenario. The Mission Context block discussed the following topics: the importance of convoy missions, some reasons UGV's were used to implement those missions, unique threats to UGV's, potential challenges to operators of UGV's in these roles, and examples of transport and remote resupply UGV's in use and/or testing. The "M" in HMT section introduced the *Sentinel* cyber-attack detection aid as the machine component of the human-machine team (HMT). It provided an overview of the *Sentinel's* purpose, design, capabilities, limitations, and its implementation for the experiment. The Anatomy of a Mission portion of the training explained the mission briefing and mission video components and discussed how they were utilized in the experiment. It also offered overview maps and a discussion of the potential mission locations. This block provided the operator familiarization training for the following items: the symbology of the mission display (**Figure 8**), the parameters of a normal UGV mission, utilization of the Mission Log Sheet **(Figure 9)** for recording speed and mission notes, and employment of the Operator Decision Tree (**Figure 7**) for calling out operator actions during the mission. The Practice Scenario was a culmination of all of the training, and it determined whether or not the operator was "ready" to move forward into the main study effort. "Ready" referred to the operator's ability to accurately perform all of the tasks of the experiment, which included recording UGV speed on the Mission Log Sheet every 30 seconds, monitoring the UGV mission, and responding to the mission scenario with

callouts from the decision tree. The practice scenario was approximately four minutes in duration, and it was designed to present the operator with both normal and abnormal mission behaviors which allowed the operator to go through a range of decision tree response callouts. The operator went through the practice run one time without any interaction from the researcher. At the conclusion of the first practice run, the researcher went back through portions of the practice mission and discussed them until the operator was familiar and comfortable with the execution of all required tasks. A second practice run was then offered to solidify the training prior to start of the main experiment. When training was complete, a five – ten minute break was taken before starting Phase 3 – Missions.

Figure 8: Mission Video Screen Shot – Annotated for Training

Mission Log Sheet

Subject Number _____ **Test ID**_____ **Date**_____

Use this Mission Log Sheet to record the UGV speed at 30 second intervals and any other notes you want to log during or after the mission.

Mission Time & UGV Speed Log	
Mission Time (min:sec)	**UGV Speed (meters/sec)**
0:30	
1:00	
1:30	
2:00	
2:30	
3:00	
3:30	
4:00	
4:30	
5:00	
5:30	
6:00	
6:30	

Mission Notes:

Figure 9: Mission Log Sheet

Phase 3 – Missions was comprised of a total of eight mission scenarios and each was 3 to 5 minutes in duration. The missions were intended to be independent of each other meaning one mission did not in any way relate to or impact another. For each mission scenario the operator was required to read the mission briefing and allowed to take notes regarding their understanding of the mission on the Mission Log Sheet. After reviewing the mission briefing, the operator was required to acknowledge understanding of the mission, review the required operator actions during the mission, observe the mission video and respond according to the context of the mission scenario. At the conclusion of each mission scenario, the operator was administered three questionnaires related to their perception of the mission. The post-mission surveys

consisted of the questionnaire regarding uncertainty and consequence, the SSI, and the NASA TLX **(Appendix III)**. The researcher always reminded the operator to respond to post-mission questionnaires in regards to the mission just completed. The process was repeated through the end of the fourth mission scenario, which was the halfway point, and the operator was then given a 5-10 minute break prior to completion of the remaining four mission scenarios. Phase 3 concluded after all eight mission scenarios and associated surveys were completed.

3.3.4 DOE: Lexicon and Scoring Approach

This section discusses some of the terminology used in the experiment and explains the process used to acquire the post-scenario data collection. The following items are discussed: 1) HMT Performance - "Score," 2) Response Time, 3) State-suspicion Index - SSI, 4) Consequence and Uncertainty, and 5) Cognitive Activation.

1) HMT Performance – "Score": The decision tree in **Figure 7** of Section 3.3.3 was used to standardize the possible response options and allowed HMT performance to be scored for a relative comparison across the sample population. Subject matter experts (SME) looked at each mission scenario, which included the mission brief and mission video, and rank ordered the possible response sequences from best to worst response. SME input was used to develop a scoring rubric for each mission scenario and a mission performance score was "awarded" based on the operator's response sequence as compared to the SME's scoring rubric. Since each mission scenario was unique and counterbalancing was used, sixteen unique scoring rubrics were created. **Table 6** was one example of a scoring rubric used during the study. In this particular example, if the operator's response sequence was 1-2-2, the mission performance score would be 80. The operator's response sequence was recorded during the experiment using the

TurningPoint interactive polling software and the ResponseCard shown in **Table 5** of Section 3.3.3. The operator verbally called out the desired response action from the decision tree during the mission, and the researcher logged the operator's decision sequence in real time using the ResponseCard. Finally, the operator was penalized (minus 10 points) for each additional decision response not associated with an actual experimental event.

Possible Operator Responses	Response Rank (best to worst)	Mission Performance Score
1 – 1	1	100
1 – 2 – 1	2	90
1 – 2 – 2	3	80
1 – 2 – 3	4	60
1 – 3 – 2	5	40
1 – 3 – 1	6	20
No Response	7	0

Table 6: Performance Score Sheet Rubric

2) Response Time – "Time": "Time" reflected the length of time in seconds required to arrive at the final decision in the operator's decision tree sequence. It was recorded during the experiment using the TurningPoint interactive polling software and the ResponseCard shown in **Table 5** for Section 3.3.3. The software logged the mission time associated with every decision tree response entered by the researcher with the ResponseCard. The response time then was the difference in mission time from the first logged decision tree response to the last logged decision tree response during the operator's decision tree sequence. In the example above, the operator's response sequence was 1-2-2. In this case the mission time was logged in the software when the researcher took the following actions: entered the input of 1 for the operator's decision to "Acknowledge" an issue, entered the input 2 for the operator's decision to "Take corrective action and/or report," and entered the input 2 for the operator's decision to "Develop solution to restore / continue." Thus the scored response time in this case would be the difference in mission times between the first logged input of 1 and the last logged input of 2.

3) State-suspicion Index (SSI): Prior to my research, there was a 20-item state suspicion index (SSI) developed by Bobko et al. that "generally" measured suspicion. I worked directly with Dr. Bobko to co-develop a 13-item contextually relevant SSI questionnaire to measure suspicion in my mission scenarios. The 13-items were questions scored on a 1 – 7 Likert scale and distributed as follows:

- 3 questions related to perception of uncertainty
- 3 questions related to perception of cognitive activation
- 3 questions related directly to suspicion
- 4 questions related to perception of malicious intent

There were two ways to score suspicion using the 13-item SSI questionnaire:

- take the average of only the three items related directly to suspicion, or
- take the average of all 13 items.

I chose the latter of the two methods to calculate total suspicion and denoted it "SSI_TOTAL. The SSI questionnaire was administered at the end of each mission scenario; therefore, an operator responded to this questionnaire eight times during the experiment. The questionnaires were always related to the mission scenario just completed.

4) Consequence and Uncertainty: In this experiment consequence referred to the operator's perceived consequence of decisions made during the mission, and uncertainty referred to the operator's perceived uncertainty about mission success. It is important to note the use of "uncertainty" in this experiment was strictly for the manipulation and influence of the operator's subjective perception of uncertainty, and it was not intended as a validated statistical measure of uncertainty. The scope of this experiment did not include statistical validation of operator uncertainty. A four-item questionnaire was developed to measure an operator's perception of

consequence and uncertainty regarding the mission scenario. The four-items were scored on a 1 – 7 Likert scale, and the four questions were distributed as follows:

- perception of consequence of decisions during the mission
- influence of perception of consequence on decision-making
- perception of uncertainty about mission success
- influence of perception of uncertainty about mission success on decision-making

The Liker score of the first consequence and uncertainty questions were used for manipulation checks in the early stages of the experimental design to ensure the context of the mission briefings achieved the desired affects. The Likert score of the first consequence question was used as an influence indicator of the operator's perception of consequence on decision-making during the experiment. The uncertainty sub-components of the SSI questionnaire were used as the primary measure of operator uncertainty during the experiment. The consequence and uncertainty questionnaire was administered at the end of each mission scenario; therefore, an operator responded to this questionnaire eight times during the experiment. The questionnaires were always related to the mission scenario just completed.

5) Cognitive Activation: Cognitive activation referred to the operator's increase cognitive activity as a result of engagement in the mission scenario. It was measured with two methods. One method used the three questions related to perception of cognitive activation from the SSI questionnaire. The other method used the NASA TLX questionnaire which measured workload in six areas: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. The operator scored each of these areas on a scale of 0 – 100 based on their experience with the scenario just completed. The NASA TLX questionnaire was used as the

primary measure of operator cognitive activation, and it was scored by averaging all six areas to get a single cognitive activation score for each scenario.

3.4 Chapter Summary

The methodology and design of experiment discussed in this chapter were the key components of this research effort. They provided the framework through which the research questions and hypotheses from Section 2.5.2 were studied. The model depicted the variables for analysis and linked the theory of suspicion to the intended application – operator detection and response to cyber-attacks on unmanned systems. The design of experiment was carefully planned in order to manipulate the experimental factors and control for many sources of variability and threats to validity. The actions taken in Sections 3.3.1 through 3.3.3 acknowledge these threats and the complexity associated with the design of a human subjects experiment. Finally, Section 3.3.4 discussed some of the terminology and measurements used in the experiment.

Chapter 4: Discussion of Analysis Results and Concerns

4.1 Chapter Overview

The purpose of this research effort was to investigate the role of operator suspicion in the detection and response to cyber-attacks on unmanned systems. Research questions and hypotheses for this effort were proposed in Section 2.5.2, and Chapter 3 discussed the methodology and design of experiment utilized to probe at each of these questions and hypotheses. The data for the research was gathered through the conduct of scenario based, human-in-loop behavioral science experiments with active duty Air Force officers as operators of an unmanned ground vehicle in a military context. In total thirty-two officers participated in the

experiment, and the data collected was summarized in **Table 7.** This chapter presents quantitative and qualitative findings associated with the data collected from the experiment to address the research questions and hypotheses. The chapter also addresses a few concerns that arose over the course of discussions about the experiment.

Phase	Variable	Description	Method	Scale	Data Points Per Person	Data Points Per Person / Experiment	Total Data Points per 32 Persons
Pre-test	Ncog	Need for cognition	Questionnaire	Likert (1-7)	18	18	576
Pre-test	Trust_Ma	Mayer Propensity to Trust	Questionnaire	Likert (1-7)	8	8	256
Pre-test	Trust_Mc	McShane Propensity to Trust	Questionnaire	Likert (1-7)	8	8	256
Pre-test	Creativity	Bobko creativity questions	Questionnaire	Likert (1-4)	2	2	64
Pre-test	IQ1	Bobko general intelligence question	Questionnaire	Likert (1-4)	1	1	32
Pre-test	GPA_U	self reported undergraduate GPA	Questionnaire	number	1	1	32
Data collected per scenario = 1 each. Test seqence = 8 scenarios.							
Post-test	Cons1	Operator perception of consequence in scenario	Questionnaire	Likert (1-7)	1	8	256
Post-test	Cons2	Influence of concequence perception on decision	Questionnaire	Likert (1-7)	1	8	256
Post-test	Unc1	Operator perception of uncertainty in scenario	Questionnaire	Likert (1-7)	1	8	256
Post-test	Unc2	Influence of uncertainty perception on decision	Questionnaire	Likert (1-7)	1	8	256
Post-test	SSI	State-suspicion index questions	Questionnaire	Likert (1-7)	13	104	3328
Post-test	TLX	NASA-TLX task workload questions	Questionnaire	0-100	6	48	1536
Post-test	Score	researcher "graded" HMT performance	"Grading"Rubiric	0-100	1	8	256
Post-test	Time	researcher "graded" operator response time	"Grading"Rubiric	continuous	1	8	256
			Total Pre-test points:		38	38	**1216**
			Total Post-test points:		25	200	**6400**

Table 7: Summary of Data Points Collected

4.2 Questions and Hypotheses

My research addressed two categories of questions and hypotheses: 1) those related to the application of suspicion theory to operator detection and response to cyber-attacks on unmanned systems and 2) those related to the theory of suspicion itself. The questions and hypotheses related to the application of suspicion theory were the primary interest, and I referred to them as Focus Questions and Hypotheses in section 2.5.2. However, the experimental design was robust and offered a unique opportunity to collect and analyze data related to the theory of suspicion itself to inform the suspicion community, and I referred to them as Response Questions

and Hypotheses in section 2.5.2. I discussed the experimental findings associated with each of these categories of questions and hypotheses in sections 4.2.1 and 4.2.2, respectively.

4.2.1 Analysis of Focus Questions and Hypotheses

The following two questions were related to the application of suspicion theory to operator detection and response to cyber-attacks on unmanned systems. They formed the primary focus of this research and were denoted as Focus Questions (FQ).

1) **Focus Question 1 (FQ-1):** How does suspicion effect human-machine team (HMT) performance?
2) **Focus Question 2 (FQ-2):** How does consequence effect the relationship between suspicion and HMT performance?

Each of these Focus Questions and the analysis of their associated Focus Hypotheses were discussed in detail in this section. First, I provided an overview of the Focus Question and discuss a summary of findings from analysis of the associated hypotheses. Then, I provided the supporting analysis of the hypotheses from which the inferences were drawn.

1) **Focus Question 1 (FQ-1):** How does suspicion effect human-machine team (HMT) performance?

   For this study a human-machine team was defined as the pairing of the operator of a cyber-physical system (e.g. unmanned ground vehicle, UGV) with a *Sentinel* cyber-attack detection aid. The performance consisted of two components, Score and Time, and each was recorded independently for each mission scenario based on the operator's response to that mission. The Score reflected the decision-making component of the performance, and Time reflected the length of time required to arrive at the decision.

Summary of FQ-1 Findings:

Operator suspicion had a significant negative impact on HMT performance (FH.1.2), and a significant positive impact on operator task response time (FH.1.4). These findings were evidenced by the operators in the experiment. The operators took longer to respond to tasks and their response sequence selections resulted in lower performance scores when they became more suspicious. Furthermore, four cyber-attack / *Sentinel* alert combinations were tested in the experiment, and the two combination without cyber-attacks had a significant negative impact on operator suspicion; whereas, the two combinations containing cyber-attacks had a significant positive impact on operator suspicion (FH.1.3). These results occurred regardless of the presence of a *Sentinel* alert, which is consistent with the finding that *Sentinel* alerts alone do not create suspicion (FH.1.1).

Analysis of Focus Hypotheses (FH) for FQ-1:

The following hypotheses were associate with FQ-1 and denoted Focus Hypotheses (FH.1). The discussion of each FH.1 addresses the theory from which it was derived, the analysis results, and offers an explanation from the results.

- **FH.1.1:** *Sentinel* alert is related to operator suspicion.

  Bobko et al. assert that environmental cues can act as triggers of state-level suspicion. Since *Sentinel* alerts can act as an environmental cue to the operator of system abnormalities, I hypothesized *Sentinel* alerts would be related to operator suspicion. However, the result of the hierarchical linear modeling (HLM) analysis yielded a *p*-value = 0.352, which was > 0.05 and, therefore, not significant. Hypothesis FH.1.1was not supported and there was no significant direct relationship between *Sentinel* alerts and

operator suspicion. A *Sentinel* alert may decrease uncertainty about the affected area of the system and prompt a more focused information search thus serving as a catalyst to the formation of suspicion; however, the *Sentinel* alert, itself, does not create operator suspicion.

- **FH.1.2:** Operator suspicion is positively related to HMT performance.

Malicious intent is a key attribute of state-level suspicion, and suspicion can lead to greater information search, more active processing of information, and consideration of multiple plausible rival hypotheses for observed behavior (P. Bobko et al., 2014). Since cyber-attacks are by nature malicious events and require consideration of multiple solutions for the observed behavior, I hypothesized operator suspicion is positively related to HMT performance meaning a suspicious operator would score better on the tasks. The result of the HLM analysis yielded *p*-value = <0.001; $\beta_{10} = -5.630$. Although the relationship between operator suspicion and HMT performance was significant (< 0.05), the direction of the relationship was negative, which meant operator suspicion reduced HMT performance on the tasks. Therefore, hypothesis FH1.2 was not supported. This relationship was depicted graphically in **Figure 10** with the equation ($Score = 89.88 - 5.63 * SSI\_Total_{i=1-7}$). The more active processing of information associated with suspicion could lead to hyper vigilance causing the operator to respond to "normal" transient system deviations. This statement was supported by significant correlations at the 0.01 level in the relationship of suspicion to number of "Acknowledgements" and suspicion to mission abort decisions. Suspicion was also found to significantly correlate at the 0.01 level to cognitive workload, and both of these findings may lead to the observed decreased operator performance.

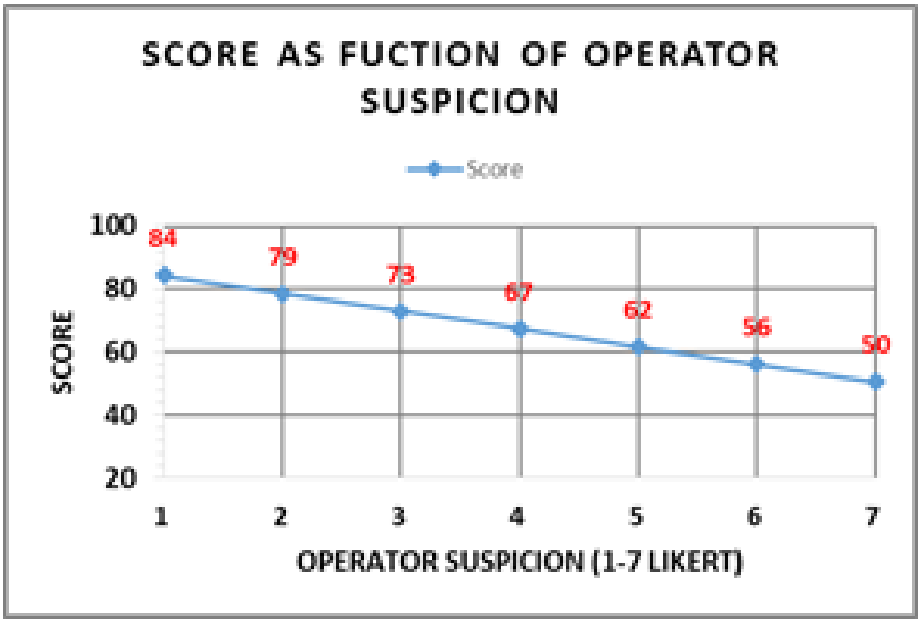


Figure 10: Graph of HMT Performance as a Function of Suspicion

- **FH.1.3:** Cyber-attack / *Sentinel* alert combinations are related to operator suspicion.

  According to Bobko et al. missing information, negative discrepancies, and distrust can lead to suspicion, which can result in greater information search, more active processing of information, and consideration of multiple plausible rival hypotheses for the observed behavior. Therefore, I hypothesized various combinations of cyber-attacks and *Sentinel* alerts are related to operator suspicion, and I created four sub-hypotheses (FH.1.3.a-d) to investigate this claim. I briefly discuss the analysis results associated with each sub-hypothesis, and I then address the observations obtained from the analysis as it relates to FH.1.3 in the "FH.1.3 – Overall Observations" section below.

  - **FH.1.3.a:** No cyber-attack / no *Sentinel* alert

    In scenarios where no cyber-attack was initiated and no *Sentinel* alert occurred, HLM analysis of the data yielded p-value = 0.019; $\beta_{10} = -0.301$, which indicated a significant relationship existed between this combination and operator

suspicion, and the relationship was in the negative direction. In other words, the combination of no cyber-attack and no *Sentinel* alert reduced operator suspicion.

- **FH.1.3.b:** Cyber-attack / *Sentinel* alert

In scenarios where a cyber-attack was initiated and the operator received a *Sentinel* alert that was accurate, HLM analysis for the data returned *p*-value = 0.047; $\beta_{20} = 0.255$, which indicated a significant relationship existed between this combination and operator suspicion, and the relationship was in the positive direction. In other words, the combination of cyber-attack and accurate *Sentinel* alert increased operator suspicion.

- **FH.1.3.c:** No cyber-attack / *Sentinel* alert (False positive, F +)

In scenarios where no cyber-attack was initiated but a *Sentinel* alert was received by the operator, HLM analysis of the data yielded *p*-value = 0.002; $\beta_{10} = -0.394$, which indicated a significant relationship existed between this combination and operator suspicion; however, the relationship was in the negative direction. In other words, the combination of no cyber-attack with a *Sentinel* alert reduced operator suspicion. This case was referred to as a F + *Sentinel* error, because the *Sentinel* led the operator to believe a cyber-attack occurred when, in fact, it had not.

- **FH.1.3.d:** Cyber-attack / no *Sentinel* alert (False negative, F -)

In scenarios where a cyber-attack was initiated but a no *Sentinel* alert was received by the operator, HLM analysis of the data returned *p*-value = 0.001; $\beta_{20} = 0.440$, which indicated a significant relationship existed between this

combination and operator suspicion, and the relationship was in the positive direction. In other words, the combination of cyber-attack with no *Sentinel* alert increased operator suspicion. This case was referred to as a F - *Sentinel* error, because the *Sentinel* failed to alert the operator to a cyber-attack when, in fact, a cyber-attack occurred.

FH.1.3 - Overall observations: The relationship between the combinations of cyber-attack / Sentinel alert and operator suspicion were shown in **Figure 11**. The mean and standard deviations were also shown in the figure and corroborate the HLM results.

Influence of Cyber-attack (CA) / Sentinel Alert (SA) Combinations on *SUSPICION*				
Combinations (N = 64 each)	HLM: Suspicion		Descriptives: Suspicion	
	p-value	Coeficient	Mean	Std. Dev
a) No CA / No SA [NaNa]	0.019	-0.301	3.966	0.958
b) CA / SA Correct [Se_Cor]	0.047	0.255	4.383	0.969
c) No CA / SA (F+) [Se_T1E]	0.002	-0.394	3.897	0.962
d) CA / No SA (F-) [Se_T2E]	0.001	0.440	4.522	1.005

CA = cyber-attack
SA = *Sentinel* alert

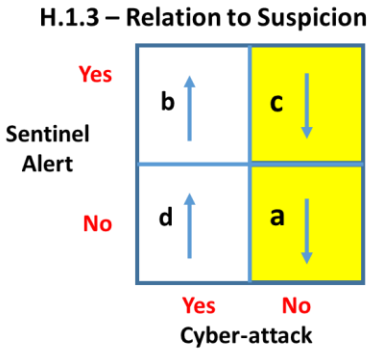


Figure 11: Summary of Attack / Alert Combination Data on Suspicion

I've made three observations from this data analysis: 1) Combinations resulting in decreased operator suspicion, 2) Combinations resulting in increased operator suspicion, and 3) Effect of *Sentinel* errors on operator suspicion.

1) Combinations resulting in decreased operator suspicion: Operator suspicion decreased when no cyber-attack occurred (a & c) regardless of receiving a *Sentinel* alert.

2) Combinations resulting in increased operator suspicion: Operator suspicion increased when a cyber-attack occurred, and the operator noticed it (b & d) regardless of receiving a *Sentinel* alert.

3) Effect of *Sentinel* errors on operator suspicion: The *Sentinel* cyber-attack detection aid was implemented to exhibit both False positive (F +) error (send alert when no attack occurred) and False negative (F –) error (no alert sent when attack occurred) types. There were 64 total instances of each error type and those cases were counterbalanced, which meant each error type resulted in 32 Low consequence cases and 32 High consequence cases. Because the analysis of *Sentinel* error types represented the mean result over the range of possible consequences, meaningful observations can be made about the overall effect of each error type. The analysis presented in **Figure 11** indicates a preference towards *Sentinel* F + errors. This preference towards F + errors was evidenced further by the data in **Table 8** which considered the effect of Cyber-Attack / *Sentinel* Alert combinations on Score (performance) and Time, as well as Suspicion. Section 4.2.1.2 provides a detailed discussion of *Sentinel* errors.

Influence of Cyber-attack (CA) / Sentinel Alert (SA) Combinations								
Combinations (N = 64 each)	HLM: Suspicion		Descriptives: Suspicion		Descriptives: Score		Descriptives: Time	
	p-value	Coeficient	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
a) No CA / No SA [NaNa]	0.019	-0.301	3.966	0.958	94.840	12.083	1.000	2.563
b) CA / SA Correct [Se_Cor]	0.047	0.255	4.383	0.969	90.780	18.109	14.910	16.447
c) No CA / SA (F+) [Se_T1E]	0.002	-0.394	3.897	0.962	92.660	15.659	4.610	3.659
d) CA / No SA (F-) [Se_T2E]	0.001	0.440	4.522	1.005	81.250	27.746	13.480	12.794

Table 8: Summary of Attack / Alert Combination Data on Suspicion, Score & Time

– **FH.1.4:** Operator suspicion is positively related to operator response time.

According to Bobko et al. suspicion leads to greater information search and more active processing of information resulting in consideration of multiple plausible rival hypotheses for the observed behavior, all of which can lead to suspended judgement. Therefore, I hypothesized operator suspicion is positively related to operator response time meaning that higher suspicion would result in longer task response time and vice a versa. The result of the

HLM analysis yielded *p-value* = *<0.001;* $\beta_{10} = 6.748$. The relationship between operator suspicion and task response Time was significant (< 0.05), and the direction of the relationship was positive, which meant operator suspicion increased task response time. Therefore, hypothesis FH1.4 was supported. Increased cognitive workload due to suspicion may lead to slower (longer) response times. This relationship was depicted graphically in **Figure 12** with the equation ($Time = 8.5 + 6.748 * SSI\_Total_{i=1-7}$).

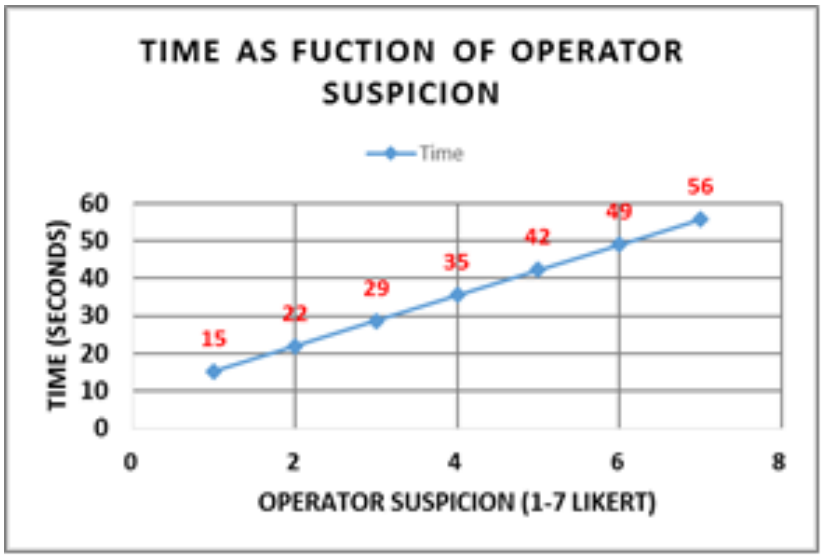


Figure 12: Graph of Operator Response Time as a Function of Suspicion

2) **Focus Question 2 (FQ-2):** How does consequence effect the relationship between suspicion and HMT performance?

For this study consequence was a two-level factor rated as either Low or High. The factor "consequence" was manipulated through the context of the mission scenario in order to create the Low or High perception of consequence within the operator. For instance, one Low consequence mission scenario was a training mission in the United States; whereas, one High consequence mission scenario was an operational mission in an undisclosed Middle-

Eastern country. The operator's perception of the consequence was measured via the post-mission scenario questionnaires in **Appendix III**.

Summary of FQ-2 Findings:

The operator's perception of the consequence associated with the mission scenario was significant and positively influenced the relationship between his/her suspicion and task response time (FH.2.2). These findings were evidenced by the operators in the experiment. The operators were more suspicious and took longer to respond to tasks when they perceived the consequence of their decisions within the mission scenario to be High. They were less suspicious and took less time to respond to tasks when they perceived the consequence of their decisions within the mission scenario to be Low. The operator's perception of the consequence associated with the mission scenario was not found to be significant and did not influenced the relationship between his/her suspicion and HMT performance (FH.2.1).

Analysis of Focus Hypotheses (FH) for FQ-2:

The following hypotheses were associate with FQ-2 and denoted Focus Hypotheses (FH.2). The discussion of each FH.2 addresses the theory from which it was derived, the analysis results, and offers an explanation from the results.

- **FH.2.1:** Consequence alters the direction or strength of the relationship between operator suspicion and HMT performance.

  As individuals become more suspicious, cognitive load will increase and rises in fear and anxiety may be experienced resulting in a decrease in processing speed and working memory (P. Bobko et al., 2014). Since a perceived elevated consequence level can lead to rises in fear and anxiety, I proposed hypothesis FH.2.1. However, the result of the

HLM analysis yielded *p*-value = 0.602, which was > 0.05 and, therefore, not significant. Hypothesis FH.2.1 was not supported and consequence did not significantly alter the direction or strength of the relationship between operator suspicion and HMT performance.

- **FH.2.2:** Consequence alters the direction or strength of the relationship between operator suspicion and task response time.

  As individuals become more suspicious, cognitive load will increase and rises in fear and anxiety may be experienced resulting in a decrease in processing speed and working memory (P. Bobko et al., 2014). Since a perceived elevated consequence level can lead to rises in fear and anxiety, I proposed hypothesis FH.2.2. The result of the HLM analysis yielded *p*-value = 0.004; $\beta_{30} = 1.18$, so it was significant and in a positive direction. Hypothesis FH.2.2 was supported. Consequence altered the direction or strength of the relationship between operator suspension and task response time such that a perceived increase in contextual consequence resulted in an increase in task response time and vice a versa. This relationship was shown graphically in **Figure 13.**

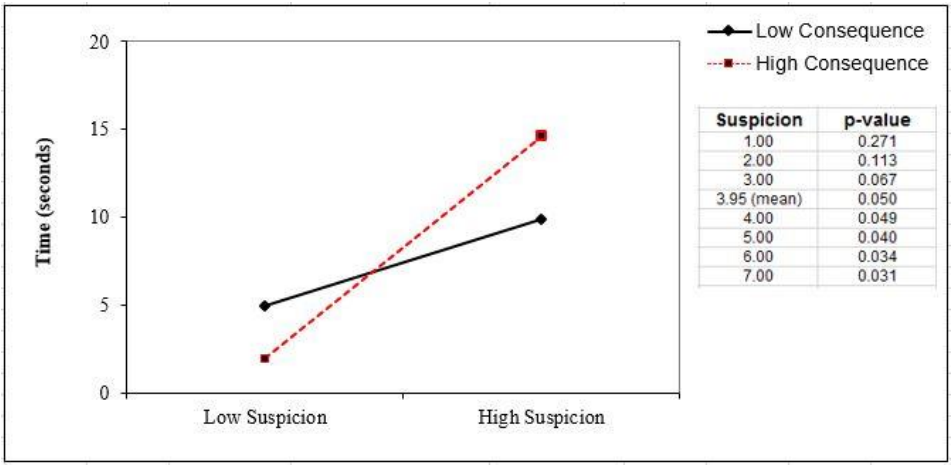


Figure 13: Graph of Suspicion and Time relative to Consequence

The plot and embedded data table shown in Figure 13 were generated using an Excel worksheet obtained from **www.jeremydawson.co.uk/slopes.htm**. The worksheet generated the two-way interaction effects for the unstandardized variables of suspicion and consequence as they relate to time, and it was used to perform a simple slope analysis to help interpret the graph. The embedded data table was computed from the slope analysis, and the following was found to be true. When suspicion was less than the mean of 3.95, consequence followed the black solid line with respect to "Time," and it did not significantly alter task response time. When suspicion was greater than or equal to the mean of 3.95, consequence followed the red dashed line with respect to "Time," and it significantly altered task response time.

4.2.1.1 Concern: Extensibility of the Dataset

When considering the analysis results of an experiment, it is important to understand the extent to which inferences can be obtained from the data to avoid over-generalization / extension of the analysis results from the dataset. I use the term extensibility to refer to this attribute of the dataset, and I ascribe two characteristics to it: 1) measurement validity and 2) range of performance of the analysis results.

1) Measurement validity: Measurement validity was discussed in Section 3.3.1: DoE: Threats to Validity as one of the threats to experimental validity, and it was a key consideration during the design of this experiment. Bobko et al.'s theory of suspicion was the construct of interest for my research, and I specifically designed to experiment to manipulate and measure its elements. The original twenty-item state suspicion index (SSI) was developed by Bobko et al. to "generally" measure the elements of suspicion, and I co-developed a contextually relevant thirteen-item SSI measure with Dr. Bobko specifically for my research. The new thirteen-item SSI measure received a Cronbach's alpha (reliability rating) of 0.881.

2) Range of Performance: The range of performance varied depending on the outcome variable of interests. Table 7: Summary of Data Points Collected contains a column titled "Scale," and it represents the possible range of performance for each of the experimental variables. When considering actual range of performance obtained from the sample, Table 8: Summary of Attack / Alert Combination Data on Suspicion, Score, and Time provides an indication of actual performance for the three main variable of Suspicion, Score, and Time. Each of those variables has a Descriptive Statistics column in the table that shows their respective Means and Standard Deviations. Taken together those actual data points give an indications of the range of performance for each of those variables over the sample.

4.2.1.2 Concern: Analysis of *Sentinel* Errors (F + and F -)

Context and consequence of decisions must be considered when examining detection errors because an individual's tolerance for one type of detection error over another is largely dependent on those two factors. I will **1)** generally define two types of detection errors and give simple examples of each, **2)** use two examples to illustrate how context and consequence can influence one's tolerance for each detection error type, and **3)** define detection error types in the context of my experiment and discuss how the analysis of these two detection error types was addressed and the inferences drawn from it.

1) False positives (F+) and false negatives (F-) are two types of detection errors typically associated with detection "systems". Generally speaking, a stimulus and a response are involved, and the type of error is related to the accuracy of the detection system in assessing and responding to the occurrence of the stimulus. A F+ detection error occurs when a response is received without the presence of the stimulus. This instance is sometimes referred to as a false

alarm. If a fire alarm were to go off when no fire or smoke was present, it would be an example of a F+ or false alarm. A F- detection error occurs when a stimulus is present but the detection system does not recognize it and send a response. This instance is also referred to as a missed detection. If a building was on fire but the fire alarm did not go off, it would be an example of a F- or missed detection. In both cases a fire alarm was the detection system, and it failed to recognize and respond appropriately to the fire stimulus.

2) I will use two different settings to illustrate the importance of context and consequence in determining tolerance to detection errors. In a cancer treatment facility doctors are more tolerant of F+ than of F- detection errors and for good reason. In this context, a F+ test result would lead a doctor and patient to believe cancer is present in the body when there really is no cancer. This diagnosis would probably lead to patient anxiety and follow-up tests, but the patient would likely find out there is no cancer and be relieved. On the other hand, a F- test result would lead a doctor and patient to believe everything is normal when cancer is actually present. This misdiagnosis creates a false sense of normalcy and could result in a treatable stage 1 cancer going undetected and growing into a non-treatable stage 4 cancer costing the person their life. The high consequence of F- detection errors in cancer testing drives a willingness to accept a degree of F+ detection errors. Alternatively, personnel in the office of a 9-1-1 dispatch center may be more tolerant of F- than of F+ detection errors. In this context, a F- detection error may mean someone with a legitimate need for assistance is not perceived as such and does not get help. A F+ detection error in this context would result in first responders being dispatched when actually not needed. Considering the high volume of 9-1-1 calls through the dispatch center, responding to multiple F+ detection errors would be very costly, and it would tie up resources that may be needed to address legitimate needs for help elsewhere.

3) In my experiment, the *Sentinel* was the cyber-attack detection aid, and it was designed to provide the operator an alert response when a cyber-attack stimulus occurred. A F+ *Sentinel* error occurred when there was no cyber-attack but the *Sentinel* sent a cyber-attack alert response to the operator. A F- *Sentinel* error ensued when a cyber-attack occurred, but the *Sentinel* did not detect it and send a cyber-attack alert response to the operator. The *Sentinel* cyber-attack detection aid was implemented to exhibit both F+ and F– detection errors. There were 64 total instances of each detection error type and two levels of contextual consequence (Low or High) possible in the experiment. The 64 total instances were counterbalanced, which meant each detection error type resulted in 32 Low consequence cases and 32 High consequence cases. Because the analysis of *Sentinel* detection error types represented the mean result over the range of possible Low / High consequences, the "necessary conditions" of context and consequence were met through the experimental design and meaningful observations were made about the overall effect of each detection error type as it related to HMT performance. The analysis presented in **Figure 11** of section 4.2.1 indicated a tolerance in HMT performance towards *Sentinel* F+ detection errors. This tolerance towards F+ detection errors was further evidenced by the data in **Table 8** of section 4.2.1 which considered the effect of Cyber-Attack / *Sentinel* Alert combinations on Score and Time – two HMT performance measures – as well as Suspicion. Although this finding seems counterintuitive, it actually follows the theory of suspicion and other findings in this experiment. The *Sentinel* alert itself did not create suspicion, but the F + *Sentinel* alert served as a catalyst for greater information search. So, when the alert was received, the operator immediately started searching for confirmatory information. The implementation of a simple cyber-attack vector in the experiment made it easily detectable via confirmatory information readily available on the mission video. Therefore, the operator could

quickly determine the alert to be False and make an appropriate decision response. Thus, the operator's response time was low and the performance Score was high.

Since the effect of cyber-attack & *Sentinel* alert combinations have huge potential implications to HMT design, a more detailed analysis and discussion was warranted. **Table 9** contains a detailed frequency count of HMT actions to each combination (a-d) in four functional areas: Operator Response, Suspicion, HMT Performance, and Response Time. All operators in the experiment responded to the mission scenarios using the decision tree in **Figure 7** of Section 3.3.3, and their response options were summarized (0-6) in **Table 9**. The data in **Table 9** for HMT actions in combinations (a) & (b), which were combinations in which the operator and *Sentinel* agree, represent expected behaviors and were not discussed in further detail. The more interesting results in **Table 9** were for HMT actions in combinations (c) & (d) which represented F + and F – scenarios, respectively.

Functional Areas	FH.1.3 a-d: *COMBINATIONS* of cyber-attack & *Sentinel* alert			
	a) No Cyber-attack & No *Sentinel* Alert	**b)** Cyber-attack & *Sentinel* Alert **Correct**	**c)** No Cyber-attack & *Sentinel* Alert **(False +)**	**d)** Cyber-attack & No *Sentinel* Alert **(False -)**
Operator Response	Frequency (N=64)	Frequency (N=64)	Frequency (N=64)	Frequency (N=64)
0 - No response	51	-	1	1
1 - Continue Mission	4	2	46	6
2 - Take action; *Sentinel* fix; continue	5	54	11	14
3 - Take action; Operator fix; continue	4	5	6	38
4 - Take action; Call backup; continue	-	-	-	2
5 - Abort; recovery; backup	-	2	-	2
6 - Abort; recovery; no backup	-	1	-	1
Suspicion (SSI Total range of 1-7) *	Frequency (N=64)	Frequency (N=64)	Frequency (N=64)	Frequency (N=64)
SSI Total: 1 - 3	10	5	12	1
SSI Total: 3 - 5	43	40	41	40
SSI Total: 5 - 7	11	19	11	23
* Higher = more suspicious				
HMT Performance (Score range 0-100)	Frequency (N=64)	Frequency (N=64)	Frequency (N=64)	Frequency (N=64)
Score: 0 - 50	-	3	1	11
Score: 50 - 75	5	4	2	3
Score: 75 - 100	59	57	61	50
Response Time (Time range 1-60 sec)	Frequency (N=64)	Frequency (N=64)	Frequency (N=64)	Frequency (N=64)
Time: 0 - 5	60	24	43	19
Time: 5 - 10	2	16	15	18
Time: 10 - 60	2	24	6	27

Table 9: Cyber-attack / Sentinel Alert Combination Frequencies

In F + scenarios the majority of the operators (46 / 64) responded as might be readily predicted (and hoped for). When the operators received the *Sentinel* alert, it prompted information search, which acted as a catalyst for suspicion. The operator was able to quickly assess from the available system information that a cyber-attack was not in effect and decided to over-ride the *Sentinel* alert continuing the mission without taking additional action. The quick

decision resulted in a somewhat symmetric distribution of suspicion scores, high HMT performance, and fast response times (as seen in **Table 9**). Additionally, there were no "call for backup" or "abort" actions. The results for F + scenarios were quite promising from a design perspective.

In contrast, HMT actions in F - scenarios were considerably less desirable. The operators did not receive a *Sentinel* alert to prompt information search. As the cyber-attack progressed, operators in these scenarios became more suspicious, took longer to respond, and generated lower HMT performance scores. Of the operators who took action, 38 chose to develop their own solution, 2 called for backup, and 14 allowed the *Sentinel* to act (which makes little sense given the *Sentinel* did not detect the attack). Perhaps more disconcerting was the frequency of missed detections and aborts. The operators completely missed the cyber-attack 7 times and aborted the mission 3 times. Overall, the false negative results were "alarming."

Given this analysis I'd recommend the developer of a cyber-attack detection aid focus their efforts on reducing the number of F- detection errors made by the cyber-attack detection aid since F- detection errors produce more costly HMT performance results in terms of operator detection and response to cyber-attacks against unmanned vehicle systems.

4.2.2 Analysis of Response Questions and Hypotheses

The following two questions were related directly to the theory of suspicion and associated propositions as proposed in (P. Bobko et al., 2014). Although secondary to my main research focus, these questions were important to the suspicion community, and my experimental designed allowed for the collection and analysis of data to provide insightful responses to the

community. These questions were denoted Response Questions (RQ), and **Appendix II** contains the pre-experiment surveys utilized to obtain the data.

1) **Response Question 1 (RQ-1):** What is the relationship between general trait-level attributes and operator suspicion?
2) **Response Question 2 (RQ-2):** How does perception of consequence affect operator suspicion?

Each of these Response Questions and the analysis of their associated Response Hypotheses were discussed in detail in this section. First, I provided an overview of the Response Question and discuss a summary of findings from analysis of the associated hypotheses. Then, I provided the supporting analysis of the hypotheses from which the inferences were drawn.

1) **Response Question 1 (RQ-1):** What is the relationship between general trait-level attributes and operator suspicion?

   Many traits potentially effect formation of suspicion; however, Bobko et al. discussed creativity, cognitive ability, need for cognition, and propensity to trust as key factors believed to be related to one's capacity to become suspicious. The experimental design allowed for the collection and analysis of data to provide novel insights concerning these propositions. The trait-level data was collected from each operator using the pre-test questionnaires found in **Appendix II**.

   Summary of RQ-1 Findings:

   Of the four individual trait-level attributes assessed in the experiment, creativity was the only individual trait to show a significant relationship to operator suspicion (RH.1.1). Operators with higher measured creativity reflected higher suspicion scores. The other three trait-level

attributes of cognitive ability, need for cognition, and propensity to trust did not show a significant relationship to operator suspicion (RH.1.2, RH.1.3, & RH.1.4). Since these pre-test questionnaires were only administered once to each operator, the sample size for the trait-level analysis was N = 32. These latter findings were not consistent with Bobko et al.'s propositions, and the inconsistency was believed to be attributed to low power of test associated with the small N for these attributes.

Analysis of Response Hypotheses (RH) for RQ-1:

The following hypotheses were associate with RQ-1 and denoted Response Hypotheses (RH.1). The discussion of each RH.1 addresses the theory from which it was derived, the analysis results, and offers an explanation from the results.

- **RH.1.1:** Creativity is positively related to operator suspicion.

  According to Bobko et al. creative people are more likely to become suspicious, which led to the formulation of hypothesis RH.1.1. Creativity was measured using a two-item self-report questionnaire utilized in some of Dr. Bobko's suspicion research activities. It had a Cronbach alpha (reliability) of 0.570, and it was administered in the pre-test phase. The HLM analysis results of this data yielded *p*-value = 0.031; $\beta_{10} = 0.333$, which indicated a significant positive relationship between operator suspicion and creativity. Therefore, RH.1.1 was supported and an increase in operator creativity resulted in an increase in the operator's capacity to become suspicious.

- **RH.1.2:** Cognitive capacity is positively related to operator suspicion.

  Bobko et al. proposed that individuals with higher cognitive capacity were more capable of becoming suspicious, because they are more capable of handling multiple plausible

hypotheses for observed behaviors while accomplishing their primary tasks. This reasoning resulted in the formulation of RH.1.2, and cognitive capacity was measured in two ways: GPA and IQ. The operator's undergraduate GPA was self-reported on the Demographic questionnaire, and a single-item self-report questionnaire from Dr. Bobko's suspicion research activity was utilized for the measure of IQ. Both of these questionnaires were administered during the pre-test phase of the experiment. HLM analysis was conducted on GPA, IQ, and GPA + IQ, and the results yielded *p*-value = 0.618, *p*-value = 0.508, and *p*-value = 0.550 + *p*-value = 0.462, respectively. The analysis results were all > 0.05; therefore, hypothesis RH.1.2 was not supported, and cognitive capacity did not exhibit a significant relationship to suspicion.

- **RH.1.3:** Propensity to trust is negatively related to operator suspicion.

  According to Bobko et al. persons with a high propensity to trust were less likely to become suspicious, which resulted in the construction of hypothesis RH.1.3. I combined the questions from the trust surveys of two known researchers – Mayer and McShane – to create one sixteen-item propensity to trust questionnaire consisting of the eight original questions from each of the two known trust surveys. This consolidated questionnaire was administered during the pre-test phase of the experiment. Although the data was collected on the same questionnaire, I performed the HLM analysis on each of the known trust surveys individually. The HLM analysis for the Mayer construct yielded *p*-value = 0.351, and the HLM analysis for McShane's construct yielded *p*-value = 0.153. Neither of these propensity to trust constructs produced significant results in the experiment; therefore, hypothesis RH.1.3 was not supported. No significant relationship was found between operator suspicion and propensity to trust through this experiment. I decided to

"officially" use the data and results collected from the Mayer propensity to trust survey questions, because it is the most well-known propensity to trust measure in academia, and it had the higher Cronbach's alpha (reliability) score of 0.752.

- **RH.1.4:** Need for cognition is positively related to operator suspicion.

  According to Bobko et al. persons with a high need for cognition were more likely to become suspicious, which resulted in the construction of hypothesis RH.1.4. I administered the eighteen-item Need for Cognition questionnaire developed by Cacioppo et al. during the pre-test phase. This measurement construct had a Cronbach's alpha (reliability) score of 0.866. The HLM analysis results yielded *p*-value = 0.299; therefore, hypothesis RH.1.4 was not supported. No significant relationship was found between operator suspicion and need for cognition through this experiment.

2) **Response Question 2 (RQ-2):** How does perception of consequence affect operator suspicion?

Considering the old adage, "Perception is reality," the experimental design supported collection of data via post-mission scenario questionnaires regarding the operator's perception of the scenario-based mission consequence (Low or High). This data was assessed to determine the potential relationship between the scenario-based consequence, the operator's perception of that consequence, and the operator's suspicion and performance.

Summary of RQ-2 Findings:

The operator's perception of the consequence associated with the mission scenario was significant and explains the relationship between his/her suspicion and task response time (RH.2.2). These findings were evidenced by the operators in the experiment. The operators

were more suspicious and took longer to respond to tasks when they perceived the consequence of their decisions within the mission scenario to be High. They were less suspicious and took less time to respond to tasks when they perceived the consequence of their decisions within the mission scenario to be Low. On the other hand, the operator's perception of the consequence associated with the mission scenario was not found to be significant and did not explain the relationship between his/her suspicion and HMT performance (RH.2.1).

Analysis of Response Hypotheses (RH) for RQ-2:

The following hypotheses were associate with RQ-2 and denoted Response Hypotheses (RH.2). The discussion of each RH.2 addressed the theory from which it was derived, the analysis results, and offered an explanation from the results.

- **RH.2.1:** Operator suspicion mediates (explains) the relationship between perception of consequence (PoC) and operator performance (Score).

  Hypothesis RH.2.1 was motivated from experience and intuition and was not directly linked to the theory of suspicion. The methodology for the analysis was discussed in Section 3.2.2.2: Analysis Approach to Response Hypotheses, and it was shown graphically in **Figure 4** of that section. As indicated in the methodology, analysis for mediation was a four step process requiring three regressions steps. The results of the HLM analysis were shown in **Table 10**. Since the first HLM regression step was not significant, there was no need to continue to the next regression step. Hypothesis RH.2.1 was not supported, and suspicion did not explain the relationship between perception of consequence (PoC) and operator performance (Score).

Suspicion explains (mediates) PoC & Score			
Step	**Outcome to Predictor**	**Significant (Y / N)**	**Beta**
1	Score to Con1	**No; *p=0.141***	-1.211
2	**Done; NOT mediator**		
3			

Table 10: Mediation Analysis of Suspicion to Consequence & Score

- **RH.2.2:** Operator suspicion mediates (explains) the relationship between perception of consequence (PoC) and task response time.

Hypothesis RH.2.2 was motivated from experience and intuition and was not directly linked to the theory of suspicion. The methodology for the analysis was discussed in Section 3.2.2.2: Analysis Approach to Response Hypotheses, and it was shown graphically in **Figure 4** of that section. As indicated in the methodology, analysis for mediation was a four step process requiring three regressions steps. The results of the HLM regression analysis sequence were shown in **Table 11**. Since regression steps 1 & 2 were significant, the analysis continued on to regression step 3. Regression step 3 showed consequence was not significant, but suspicion was significant. The analysis results supported hypothesis RH.2.2. Since suspicion was significant, adding suspicion to the model reduced the Time beta from 2.426 to 0.547, which was an indicator of partial mediation between the operators' perception of the consequence and task response time.

Suspicion explains (mediates) PoC & Time			
Step	**Outcome to Predictor**	**Significant (Y / N)**	**Beta**
1	Time to Con1	Yes; *p=<0.001*	2.426
2	SSI_Total to Con1	Yes; *p=<0.001*	0.301
3	Time to Con1 + SSI_Total	Con1; No; *p=0.314*	0.547
		SSI_Total; Yes; *p=<0.001*	6.246

Table 11: Mediation Analysis of Suspicion to Consequence & Time

4.2.2.1 Concern: Testing Rare Events and Sequencing

Cyber-attacks against physical systems are considered rare events, so one may question how an experiment can accurately represent an operator's response to an event s/he may only encounter once in a career, if at all, since the experiment exposed the operator to multiple "cyber-attacks" over a short duration – a repeated measures design. This is a very common occurrence in military and commercial aviation domains as pilots must be trained to identify and respond to "rare events" such as aircraft system failures in flight. In a conference paper titled, "Test Scenarios for Rare Events," Newman and Foyle reviewed experimental studies over the past several years with the goal of developing experimental scenarios to test rare events in aviation that will produce suitable data while making efficient use of experimental facilities. They noted the similarity of rare events to vigilance studies but acknowledged the impracticality – in terms of expense and time – to place a pilot in a simulator for many trials until a "rare event" happens. They also surmised a similar concern that rare events are experimentally difficult to handle because the crux of the problem is "how do we test pilot response to rare events?" (Newman & Foyle, 2003). The cases reviewed by Newman and Foyle were all simulation based experiments, which appears to be the industry standard method of testing "rare events" in

aviation. Although they do not state explicitly the nature of the experimental designs studied as being repeated measures, the context of the discussion leads me to believe they were. Fortunately, Newman and Foyle offered recommendations to improve experimental testing of "rare events" and offered recommendations to the aviation industry. The context of their interest was situational awareness for decision-making, and they made the following recommendations pertaining to the design of experimental studies for testing "rare events": 1) develop operational scenarios, 2) model human error, 3) develop test objectives, and 4) develop objective test criteria.

Further details of their recommendations can be found in (Newman & Foyle, 2003), and I related their recommendations to my DoE.

1) Develop operational scenarios: I discussed this point at length in Section 3.3: DoE and more specifically in Section 3.3.2 DoE: Scenario Development. The context of the scenarios and the setting of the experiment were operationally realistic.

2) Model human error: Newman and Foyle are specifically referring to the Situational Awareness (SA) Error Taxonomy developed and presented in (Endsley, 1988). Newman and Foyle summarized Endsley's taxonomy in Table I of their conference paper, and I replicated it in **Table 12** below for quick reference. Each of the three Levels presented in **Table 12** were evaluated through the DoE.

Table I: Situation Awareness Error Taxonomy

Error Type	Description
Level 1 Failure to Correctly Perceive Element	
Data not available	Data are not available due to failure of the system design
Data hard to detect	Examples: inadequate lighting or resolution
Failure to observe data	Data not perceived due to omission, attentional limitations, distractions, etc.
Misinterpretation	Data misperceived because of prior expectations or misunderstood because of distraction
Cognitive failures	Caused by high workload
Level 2 Failure to Comprehend Elements	
Poor mental model	Poor mental model does not allow combining for information needed to meet goals
Incorrect mental model	Leads to incorrect assessment
Over reliance on default model	Routine expectation of the system is assumed
Other	
Level 3 Failure to Project Future State	
Poor mental model	Poor mental model does not allow for projection into future state
Over projection of current trends	Projection further into future than current data warrants
Other	

Table 12: Situational Awareness Error Taxonomy

3) Develop test objectives: Newman and Foyle suggested development of test objectives derived from the confluence of intended use and the human error taxonomy. The DoE's primary objective was to replicate the theory of suspicion in an operationally realistic way to evaluate the relationship of operator suspicion to HMT performance in detection and response to cyber-attacks against their unmanned ground vehicle mission. Detection and response were the two overarching objectives, and Section 3.3.3: DOE Operationalization discusses them in detail. Newman and Foyle stated many studies concentrated on pilot reaction; fewer examined the pilots' ability to recognize a situation; and even fewer have examined pilots' reaction to a "rare event" (Newman & Foyle, 2003). My DoE was unique in that it specifically examined operator detection and response to the rare event of a cyber-attack on an unmanned system.

4) Develop objective test criteria: Newman and Foyle also stressed the importance of employing objective metrics such as reaction time, accuracy of decision, etc. as much as possible. As

discussed in Section 3.3: DoE, I collected both subjective and objective data from each scenario, and a summary of the data collected was placed in **Table 7** of Section 4.1.

Additionally, test subject learning and memory are typical potential concerns associated with a repeated measures DoE, and these concerns could give rise to further hesitations associated with testing "rare event." As discussed in Section 3.3.1: DoE Threats to Validity, the experiment was designed to counteract these effects through counterbalancing and randomization. These procedures are "industry standard" approaches to dealing with experimental threats due to learning and memory.

Finally, the sequence in which the operator was exposed to the different mission scenarios could give rise to concern about the analysis results related to experimental tests associated with "rare events." As discussed in Section 3.3.1: DoE Threats to Validity, the sequence in which the operator was exposed to mission scenarios was randomized. I perform an autocorrelation analysis on the experiment to determine if experimental sequence was a concern. The result of the autocorrelation analysis presented in **Figure 14** confirmed sequence was not an issue for this experiment.

The DoE for my research was robust; therefore, it was feasible to believe results from this experiment addressed concerns associated with testing "rare events."

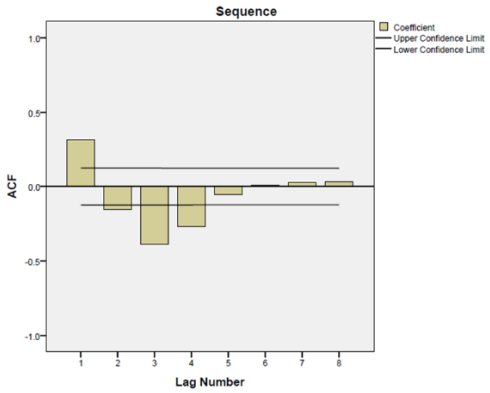


Figure 14: Autocorrelation Results for Experiment Sequence

4.3 Experiment Limitations

All experiments have limitations, and this research was no exception. First, this was novel research and a baseline for the effects of suspicion on operator response to system anomalies did not exist. Further, there was no baseline for operator performance in this environment with a *Sentinel* cyber-attack alert system. Therefore, the analysis results and inferences gained through the experiment were first looks into these areas, and the results needed to be interpreted within the confines, context, and setting of the experiment.

A second limitation associated with the experiment concerned the low sample size available to address the individual trait-level characteristics proposed by Bobko et al. for propensity to become suspicious. The experiment was conducted with a total of 32 Air Force officers serving as unmanned ground vehicle operators. Each officer experienced 8 mission scenarios which provided an N = 256 and resulted in significant experimental power for inferences about the Focus questions and hypotheses. On the other hand, the pre-test questionnaires constructed to address trait-level operator characteristics were only administered once to each operator and resulted in a sample size of N = 32 and low statistical power for trait-level analysis. Of the four individual trait-level attributes assessed in the experiment, creativity was the only individual trait to show a significant relationship to operator suspicion (RH.1.1). The other three trait-level attributes of cognitive ability, need for cognition, and propensity to trust did not show a significant relationship to operator suspicion (RH.1.2, RH.1.3, & RH.1.4).

These latter findings differed with Bobko et al.'s propositions, and the inconsistency was believed to be attributed to low power of test associated with the small N for these attributes.

A third limitation associated with the experiment was the construct of the HMT performance measurement. The HMT performance measurement was intentionally developed to consist of two components: Score and Time. Score represented the operators' performance based on their response sequence for each mission scenario, and Time represented the total amount of time in seconds it took the operator to respond to the tasks. These performance measures were taken independently of each other in order to gain insight into how each component was effected by suspicion. The results of the experiment showed operator suspicion was not related to Score, but it was related to Time. Further investigation indicated the potential for a direct negative relationship between Time and Score; therefore, a likely indirect relationship existed between suspicion and Score through Time. It may be beneficial to consider a new HMT performance measure that combines the components of Score and Time into one factor and examine how suspicion then effects it.

Another potential limitation of the experiment involves the selection of just one cyber-attack vector. The primary objective of the research was to explore the relationship between operator suspicion and detection / response to cyber-attacks on unmanned systems. As such, I was careful not to make the type of cyber-attack (attack vector) a factor in the DoE. Therefore, I selected one cyber-attack vector and used it for every case involving a cyber-attack. The cyber-attack vector I chose to simulate was an attack on the throttle control of the unmanned ground vehicle. This attack vector was non-complex and more easily detectable than other attack vectors such as a more complex attack on the navigation system and display. It was possible the complexity and manifestation of the attack vector could influence the outcome of the results.

The last limitation I'll discuss involves generalizability of the test results to other situations and / or communities. According to Sackett, generalizability is a function of

methodology, not results, and the degree to which outcomes can be generalized is either built into or out of the experimental design. The actions taken in Sections 3.3.1 through 3.3.3 acknowledge many threats to experimental validity and attest to the complexity associated with the design of a human subjects experiment. However, despite these design efforts, limitations still exist regarding generalizability. The focus of the experiment was on military operators of unmanned vehicles, because the motivating issue was initially observed in Air Force operators of remotely piloted aircraft systems (RPAS). However, due to the high mission demand and operational tempo of these assets, it was not practical to employ RPAS and RPAS operators. Therefore, an unmanned ground vehicle system was utilized to represent a more generalized unmanned vehicle, and Air Force members at the Air Force Institute of Technology (AFIT) were used as the operators. I believe the experimental system and operators were representative surrogates; however, others may disagree. For instance, the Army tends to utilize enlisted military personnel for RPAS operations, so the Army may not view Air Force officer as a representative sample for them. Therefore, it was feasible to believe results from this experiment would generalize to operators of unmanned systems in a military setting; however, the uniqueness of individual service operations must be considered. It may not be wise to attempt generalization beyond a military setting and across operational communities.

Chapter 5: Summary and Conclusions

5.1 Chapter Overview

This chapter will summarize my dissertation. First, it will review the purpose and scope of the dissertation. Then, it will discuss the research contributions. Finally, it will conclude with a discussion of future work to extend the impact and relevance of this research.

5.2 Review of Purpose and Scope

The purpose of my research was to study the relationship of operator suspicion to his/her detection and response to cyber-attacks on unmanned system operations. The scope of the document was as follows. Chapter 1 provided an introduction to the topic and motivation for the dissertation. Chapter 2 introduced the theory of suspicions, the theoretical model for analysis, and the research questions and hypotheses. Chapter 3 provided a description of the analysis methodology and the design of experiment developed to execute the research objectives. Chapter 4 delivered a discussion of the analysis results and some associate concerns.

This research was motivated by findings from a 2013-14 experiment with Air Force operators of remotely piloted aircraft systems (RPAS). During this earlier experiment, mission performance was severely degraded or missions were aborted because the operators were unable to detect simulated cyber-attacks – without the assistance of a *Sentinel* automated cyber-attack detection aid, – didn't know how to respond to *Sentinel* alerts if received, and never suspected malicious intent as a cause for their system's anomalies. These issues highlighted the need for my research and prompted the literature review which led to the application of suspicion theory to operator detection and response to cyber-attacks on unmanned systems.

The context and setting where the initial observations occurred set the boundaries for this dissertation and aided in defining the scope of the experiment. It was impractical – due to high mission demand and operational tempo – to conduct my research using actual Air Force RPAS and RPAS operators; however, my research needed to emulate this context and setting as closely as feasible to explore solutions to the observed issues and work towards generalizability. Thus, the scope was limited to a functionally representative unmanned vehicle system, operationally

relevant mission scenarios and setting, and Air Force officers from Wright-Patterson AFB where the experiments were conducted.

I further constrained the scope through the selection of experimental variables. Those variables included the components of Bobko et al.'s suspicion theory (uncertainty, cognitive activation, and perception of malicious intent), mission consequence, one cyber-attack vector (throttle control), and a Sentinel cyber-attack detection aid. The experimental design consisted of two interactive components, which comprised the mission scenario: the mission briefing and the mission video. Each component of the mission scenario was designed to manipulate these factors either Low / High or On / Off. I used an unmanned ground vehicle (UGV) to run the carefully constructed missions and recorded them live with screen capture software (CamStudio and TinyTake). Video editing software (Filmora Video Editor) was used to edit the mission videos to create the desired experimental effect. These mission videos provided a controlled and repeatable platform for operator interaction during the experiment.

The experimental setting and tasks were also scoped to functionally represent RPAS operations. The setting was an office environment with a computer and monitor for running the training and mission scenarios. The operators were required to perform multiple tasks involving monitoring the mission video feed, recording mission parameters on a mission log sheet, and responding to events during the mission. These tasks and interactions were similar to those of RPAS operators and allowed for collection of operator trait and performance data for analysis.

5.3 Research Contributions

Cyber-attacks against cyber-physical systems are serious and emergent threats with potentially catastrophic impacts, and the topic has garnered considerable interest. Much research

is being done to address the physical security aspects of cyber-physical systems; however, research addressing the human dimension of cyber-attack response from an operator and operational perspective is sparse. My research was a unique probe into the factors affecting operator resilience to cyber-attacks, which are situations characterized by uncertainty and malicious intent.

The variability of individual operators make it improbable to grasp the full range of factors contributing to operator performance in every situation; however, the literature review provided a starting point to aid in understanding operator performance in situations involving malicious intent (i.e. a cyber-attack), and the concept of suspicion was believed to be a key factor in operator response to cyber-attacks. My research effort explored this human dimension through scenario based, human-in-the loop behavioral science experiments with Air Force personnel. It included both abstract and empirical assessments of the application of suspicion theory to operator detection and response to cyber-attacks against an unmanned vehicle system, and it took a systems-oriented approach to the problem by incorporating a human-machine team (HMT) in the response. The HMT was defined as an operator (human) and a *Sentinel* (an automated hardware / software cyber-attack detection aid).

My research was a novel approach toward addressing the issue of cyber-attacks against cyber-physical systems such as unmanned vehicle systems. The contributions of my research were numerous and were grouped into three categories: experimental design, research findings, and research implications.

Experimental design:

Since my research effort was the first to apply the theory of suspicion to operator detection and response to cyber-attacks on unmanned systems, I designed the entirety of the

experiment. Some of the contributions included the development of a theory based model, mission briefings, a test setup, and metrics. Each are discussed below.

- **Theory based model:** The first step was developing the theory-based model presented in **Figure 3** for empirically testing operator suspicion. The model documented the relationship between the components of the suspicion theory and the observable / testable elements which linked operator characteristics and responses to the theory of suspicion. The model provided a roadmap for measurement selection / development and the associated analysis.
- **Mission briefings:** The next major task and contribution was the development of the mission briefing in **Appendix I**. The mission briefings operationalized the factors of suspicion theory into realistic operational mission contexts for an unmanned ground vehicle system. Every aspect of the mission briefing was specifically designed to implement a component of theory in a way that would manipulate the operator to perceive a desired level (Low or High) of that component. These mission briefings underwent manipulation checks and a pilot study to ensure they achieved the desired effect. They were discussed in Section 3.3.2: DoE – Scenario Development and can serve as a template for anyone needing to operationalize the theory of suspicion (and possibility other related theories).
- **Test setup:** Another contribution was the development of a cyber-physical system test bed for experimenting with cyber-attacks in an operationally relevant way and recording those missions for editing, play back, and operator interaction. Section 3.3.3: DoE – Operationalization discussed the details of the test setup, and **Figures 5 & 6** depicted the test setup used for my research. This was a versatile test set up, and it can be used to

implement other types of cyber-attacks and evaluate the role of operator suspicion to different cyber threats.

- **Metric:** Lastly, metric development was a significant contribution. Prior to my research, there was a 20-item state suspicion index (SSI) developed by Bobko et al. that "generally" measured suspicion. I worked directly with Dr. Bobko to co-develop a 13-item contextually relevant SSI to measure suspicion in mission scenarios. This new SSI metric received a Cronbach alpha (reliability) score of 0.881, and it has a higher reliability than the original 20-item SSI metric. The 13-item SSI metric can be used for other operational mission focused research efforts, and it can serve as a model for how to tailor the original general SSI metric to measure suspicion in a specific context.

Research Finding:

Considering the uniqueness of my research in addressing the issue of operator response to cyber-attacks on unmanned systems, I believe many of my findings contribute to research in the areas of suspicion (in general) and human response to cyber-attacks (specifically). My research consisted of four questions which were discussed at length in Sections 4.2.1 & 4.2.2. I restated each question below with a summary of the findings associated with it.

1) **Focus Question 1 (FQ-1):** How does suspicion effect human-machine team (HMT) performance?
   - Sentinel alerts alone did not create operator suspicion (FH.1.1)
   - Increases in operator suspicion negatively impacted important HMT performance metrics.

     Experimental evidence: lower performance scores, increased mission aborts, and increased operator responses (FH.1.2).

- Operator suspicion was influenced by Cyber-Attack / *Sentinel* Alert combinations

  Four cyber-attack / *Sentinel* alert combinations were tested in the experiment. The two combination without cyber-attacks had a significant negative impact on operator suspicion; whereas, the two combinations containing cyber-attacks had a significant positive impact on operator suspicion. These results occurred regardless of the presence of a *Sentinel* alert (FH.1.3).

- Analysis indicated a tolerance in HMT performance towards *Sentinel* F+ detection errors.

  The analysis presented in **Figure 11** and **Table 8** of section 4.2.1 and **Table 9** of Section 4.2.1.2 support this finding. Although this finding seemed counterintuitive, it followed the theory of suspicion and other findings in this experiment as discussed in Section 4.2.1.2.

- Increases in operator suspicion increased operator task response time (FH.1.4).

2) **Focus Question 2 (FQ-2):** How does consequence effect the relationship between suspicion and HMT performance?
   - Consequence did not influenced the relationship between operator suspicion and HMT performance (FH.2.1).
   - Consequence strengthened the relationship between operator suspicion and task response time (FH.2.2).
3) **Response Question 1 (RQ-1):** What is the relationship between general trait-level attributes and operator suspicion?
   - Creativity was the only individual trait tested to show a significant relationship to operator suspicion (RH.1.1).

- Trait-level attributes of cognitive ability, need for cognition, and propensity to trust did not show a significant relationship to operator suspicion (RH.1.2, RH.1.3, & RH.1.4). Since the pre-test questionnaires were only administered once to each operator, the sample size for the trait-level analysis was N = 32. These latter findings were not consistent with Bobko et al.'s propositions, and the inconsistency was believed to be attributed to low power of test associated with the small N for these attributes.

4) **Response Question 2 (RQ-2):** How does perception of consequence affect operator suspicion?
   - The operator's perception of the consequence was not significant and did not explain the relationship between his/her suspicion and HMT performance (RH.2.1).
   - The operator's perception of the consequence was significant and explained the relationship between his/her suspicion and task response time (RH.2.2).

     The operators were more suspicious and took longer to respond to tasks when they perceived the consequence of their decisions within the mission scenario to be High. They were less suspicious and took less time to respond to tasks when they perceived the consequence of their decisions within the mission scenario to be Low.

Research Implications:

The contributions of both the experimental design and the research findings could have far reaching impacts to the study of operator response to cyber-attacks against unmanned vehicle systems and the general study of suspicion. I believe the experimental methodology and approach developed to test a "soft" topic like suspicion has potential to benefit other behavioral science experiments. Additionally, the framework has been developed to allow for more extensive studies of human response to cyber-attacks such as looking at operator response to

other types of cyber-attacks and different operator response protocols. Much research has been accomplished in the area of cyber-attack detection aids (e.g. *Sentinel*), but the findings show that *Sentinel* alerts alone did not create operator suspicion. Instead, alerts served as a catalyst for wider information search which could lead to formation of operator suspicion if confirmatory information is not readily available to the operator for assessment of the situation. Operator suspicion is essentially a state of suspended or postponed decision-making (judgement) and remaining in a state of suspicion was demonstrated to have a negative impact on HMT performance. Therefore, it was desirable to move quickly from a state of suspicion to a decision. The presence of a *Sentinel* alert prompted a focused information search. When confirmatory data was readily available and returned from the focused information search, the operator was better able to transition through state-suspicion to a decision quickly resulting in better HMT cyber-attack detection and response performance.

The strong influence in cyber-attack and *Sentinel* alert combinations highlighted the important influence degree of automation can play in responding to cyber-attacks and how the HMT design can influence suspicion, which in turn, influences HMT performance. As system developers consider the balance of F + and F - errors in the design of cyber-attack detection aids, the results of this experiment suggest that erring on the side of F + and ensuring confirmatory information was readily available to the operator had more desirable HMT performance outcomes..

Finally, there was not a direct relationship made between suspicion and HMT performance in the experiment; however, a direct relationship between response time and HMT performance was noted. This is potentially important because of the direct relationship between suspicion and time. It is highly possible that suspicion has a significant relationship to HMT

performance through the time variable and this may be seen through an enhanced HMT performance metric that includes a function of time.

5.4 Future Work

This study identified several other opportunities for future research. Some of the prospects for follow-on research activities are provided in the list below with a brief description of the effort.

- Study the effect of operator suspicion to performance without operator knowledge and influence of a cyber-attack alert system (e.g. *Sentinel*).

  My research effort assumed the presence of a *Sentinel* in all scenarios; however, cyber-attack detection aids of this kind are not currently operational in Air Force unmanned vehicle systems. Therefore, research should be conducted to baseline the effect of just suspicion (no *Sentinel*) on operator detection and response to cyber-attacks against unmanned vehicle systems. The "suspicion only" baseline for operator performance could then be compared to the HMT responses and performance of this study.

- Study the effect of suspicion and consequence on operator performance with an enhanced performance measure that includes time as a factor.

  My research did not show a direct relationship between suspicion and HMT performance; however, a direct relationship between response time and HMT performance was noted. Since there is a direct relationship between suspicion and time, it is highly possible that suspicion has a significant relationship to HMT performance through the time variable and this may be seen through an enhanced HMT performance metric that includes a function of time.

- Study the relationship between type of attack and operator suspicion.

  I intentionally did not make the type of cyber-attack a factor for analysis in my experimental design, and I implemented a relatively easy to detect "throttle control" cyber-attack. Suspicion was still significant to many relationships even with this "simple" type of cyber-attack. Research should be undertaken to examine the relationship of operator suspicion to HMT performance in scenarios involving more complex or subtle cyber-attacks (e.g. navigation, camera, multi-stage, etc.).

- Study discrepancy resolution a *Sentinel* alert and human in the decision loop

  For my experiment, the *Sentinel* was the only "external" source of additional information. The operator would rely on the mission video and associated system parameters to determine whether or not to "believe" the *Sentinel* alert. In most operational scenarios, additional humans would be involved in the mission and decision loop, and their inputs would also have to be considered. Research should be conducted to determine how best to resolve a discrepancy between a *Sentinel* alert and contradictory information provided by a human who is in the decision loop.

- Study when in the mission timeline and / or checklist is it best for suspicion to occur.

  In this research, suspicion was potentially prompted by a *Sentinel* alert; however, a *Sentinel* may not be available to provide that prompt. Currently, operators use maintenance and operations checklists to trouble shoot a system anomaly. These checklists assume mechanical / software issues or operator error as the main cause of the anomaly and do not offer any prompts to consider the possibility of malicious intent (e.g. cyber-attack) as the cause. Research should explore the various placement of prompts

(beginning, middle, and end) in maintenance and operations checklists to consider the possibility of malicious intent as the cause for system anomalies.

Finally, all of the recommended additional future studies would benefit from the development of a higher fidelity simulation with more confirmatory and distracting data elements to support the study of the relationship of operator suspicion to the detection and response to cyber-attacks against unmanned vehicle systems.

5.5 Conclusions

In summary this dissertation took a novel approach toward addressing the very serious and emergent treat of cyber-attacks against cyber-physical systems such as unmanned vehicle systems. The very nature of a cyber-attack is one characterized by malicious intent, and the theory of suspicion as proposed by Bobko et al. has malicious intent as one of its primary components. Therefore, my approach considered the relationship of operator suspicion to the detection and response to cyber-attacks against unmanned vehicle systems. This research effort was the first of its kind to take a systems-oriented approach to the problem by incorporating a human-machine team (HMT) in the response and exploring the human dimension of suspicion through scenario based, human-in-the loop behavioral science experiments with Air Force personnel. The work accomplished in this area was by no means complete or exhaustive; however, it provides a firm foundation for beginning to think through the human dimensions of cyber-attacks.

References

ArduPilot. (2016). Mission Planner v1.3.31. Retrieved from http://ardupilot.org/planner/

Baron, R. M., & Kenny, D. a. (1986). The Moderator-Mediator Variable Distinction in Social The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. http://doi.org/10.1037/0022-3514.51.6.1173

Bobko, P., Barelka, A., Hirshfield, L. M., & Lyons, J. B. (2014). Invited Article: The Construct of Suspicion and How It Can Benefit Theories and Models in Organizational Science. *Journal of Business and Psychology*, 1–20.

Bobko, P., Barelka, a. J., & Hirshfield, L. M. (2014). The Construct of State-Level Suspicion: A Model and Research Agenda for Automated and Information Technology (IT) Contexts. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*(3), 489–508. http://doi.org/10.1177/0018720813497052

Buller, D., & Burgoon, J. (1996). Interpersonal Deception Theory. *Communication Theory*, *3*, 203–242.

Cacioppo, J. T. ., Petty, R. E., & Kao, C. F. (1984). The Efficient Assessment of Need for Cognition. *Journal of Personality Assessment*, *48*(3), 306–307.

Cacioppo, J. T., Petty, R. E., Feinstein, J. a., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*(2), 197–253. http://doi.org/10.1037/0033-2909.119.2.197

CamStudio. (2013). CamStudio 2.7.2 (Build r326). Retrieved from http://camstudio.org/

Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, *112*(1), 155–159. http://doi.org/10.1037/0033-2909.112.1.155

Endsley, M. R. (1988). Design and Evaluation for Situation Enhancement. In *THE HUMAN FACTORS SOCIETY-32nd ANNUAL MEETING* (pp. 97–101).

Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing Moderator and Mediator Effects in Counseling Psychology Research. *Journal of Counseling Psychology*, *51*(1), 2004. http://doi.org/10.1037/0022-0167.51.1.115

Gay, C. A., Horowitz, B., Lau, N., Leach, K., Dinsmore, M., Lewis, M., … Rush, J. (2015). *RT-115 Project Report - Part 2, Human Factors Engineering and System-Aware Cybersecurity*.

Hilton, J., Fein, S., & Miller, D. (1993). Suspicion and Dispositional Inference. *Personality and Social Psychology Bulletin*, (19), 501–512.

Horowitz, B. M., & Jones, R. A. (2015). *Smart Cyber Security Sentinels for Providing Cyber Security of Critical System Functions: Unmanned Air Vehicle Case Study*.

Horowitz, B. M., & Pierce, K. M. (2013). The Integration of Diversely Redundant Designs, Dynamic Systems Models, and State Estimation Technology to the Cyber Security of Physical Systems. *Systems Engineering*, *16*(4), 401–412. http://doi.org/10.1002/sys

Jones, R. A., & Horowitz, B. M. (2012a). A System-Aware Cyber Security Architecture. *Systems Engineering*, *15*(2), 225–240. http://doi.org/10.1002/sys

Jones, R. A., & Horowitz, B. M. (2012b). A System-Aware Cyber Security Architecture. *Systems Engineering*, *15*(2), 225–240. http://doi.org/10.1002/sys

Kirk, R. E. (1995). *Experimental Design: Procedures for the Behavioral Sciences* (3rd ed.). Brooks/Cole Publishing Company.

Lee, J. D., & See, K. a. (2004). Trust in automation: designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. http://doi.org/10.1518/hfes.46.1.50.30392

MangoApps, I. (2016). TinyTake. Retrieved from https://tinytake.com/

Mayer, J., & Mussweiler, T. (2011). Suspicious spirits, flexible minds: When distrust enhances creativity. *Journal of Personality and Social Psychology*, *101*(6), 1262–1277. http://doi.org/10.1037/a0024407

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integative Model of Organizational Trust. *Academy of Management Review*, *20*(3), 709–734.

Mcknight, D., Choudhury, V., & Kacmar, C. (2002). The impact of initial customer trust on intentions to transact with web site: A trust building model. *Journal of Strategic Information Systems*, *11*, 297–323.

NASA. (2016). NASA TLX: Task Load Index. Retrieved December 1, 2016, from https://humansystems.arc.nasa.gov/groups/tlx/

Neale, J. M., & Leibert, R. M. (1986). *Science and Behavior - An Introduction to Methods of Research* (3rd ed.). Englewood Cliffs, New Jersey: Printice-Hall.

Newman, R. L., & Foyle, D. C. (2003). Test Scenarios for Rare Events. In *Twelfth International Symposium on Aviation Psychology* (pp. 873–882). Retrieved from http://human-factors.arc.nasa.gov/ihi/hcsl/publications/Newman_AvPsyc03.pdf

Osborne, J. W. (2000). Advantages of Hierarchical Linear Modeling. *Practical Assessment, Research, and Evaluation*, *7*(1), 1–3.

Sackett, P. R., & Larson, J. (1990). Research strategies and tactics in I/O psychology. In M. D. Dunnette & L. Hough (Eds.), *Handbook of Industrial and Organizational Psychology (2nd)* (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An Integration Model of Organizational Trust: Past, Present, and Future. *Academy of Management Review*, *32*(2), 344–354.

http://doi.org/10.5465/amr.2007.24348410

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inferences*. Belmont, CA: Wadsworth Cengage Learning.

Turning Techologies, L. (2013). TurningPoint 5.3. Retrieved from www.turningtechnologies.com

Woltman, H., Feldstain, A., MacKay, C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 52–69. http://doi.org/10.2307/2095731

Wondershare. (2016). Filmora Video Editor v7.2.0. Retrieved from https://www.wondershare.net/

Appendix I – Mission Scenarios

As of August 12, 2016

- Mission ID: S1
- Mission Location: United States, Tonopah Test Range, Nevada
- Mission Briefing:
  - Description: Your Unit is preparing for deployment and must pass a Pre-Deployment Readiness Qualification mission at Tonopah Test Range, Nevada for autonomous (UGV) resupply of a supported unit. The cargo consists of ammunition, POL (petroleum, oil, lubricants), food, and water. It is a routine box (return) route resupply training mission proceeding through four waypoints at a speed of ~2 m/s with mission duration of ~4:00 minutes. Waypoint marker 2 is the resupply location, and the UGV should pause for ~15 seconds at the location to simulate offloading of supplies.
  - Threat Environment: Missions in this area have been protested due to their proximity to the Wildhorse Management Area and Area 51 (see map), but there is no intel suggesting activist actions during the mission.
  - Likelihood of Mission Success: Only 1 of the last 15 autonomous missions in this area sporadically reported interference with autonomous operations. The issues were believed to have been caused by interference from electronic combat range equipment at the nearby Nellis Air Force Range. Due to the low likelihood of activist actions and sporadic nature of past interferences, autonomous missions should proceed with an expected certainty for success.
  - Risk: If the Unit fails to pass the Pre-deployment Readiness Qualification mission, it will not be categorized as worldwide qualified. The unit will be "recycled" and sent through another training program to ensure proficiency of UGV operations, and another unit will be tasked to fill the deployment rotation.

As of August 12, 2016

- Mission ID: S2
- Mission Location: United States, Nellis Air Force Range, Nevada
- Mission Brief:
  - Description: Your Unit is participating in a Flag Exercise at Nellis Air Force Range, Nevada, and its mission is the autonomous delivery of a dual mode (manned / unmanned) ATV sized UGV with supplies to a remote operating unit. The unit is conducting a mock assault on a supposed remote terrorist leader compound as part of an integrated exercise. This is a routine out route training mission proceeding through three waypoints at a speed of ~2 m/s with mission duration of ~4:00 minutes.
  - Threat Environment: Missions in this area have been protested due to its proximity to the Wildhorse Management Area and Area 51 (see map), but there is no intel suggesting activist actions during the Flag Exercise. Range activity increases approximately 10 fold during Flag Exercises.
  - Likelihood of Mission Success: Interference with autonomous operations has been reported on as many as 10 of the last 20 autonomous missions occurring during Flag Exercises. The previous issues were believed to have been caused by interference from increased utilization of electronic combat range equipment at the Nellis Air Force Range; therefore, autonomous mission success during Flag Exercises is uncertain.
  - Risk: This is an integrated training mission with surveillance and close air support (CAS) assets supporting an assault on the mock terrorist leader compound, and there is a five hour window in which the supposed terrorist leaders are suspected to be at the compound. The remote unit is waiting on the supplies and transport to conduct the assault. Failure of your UGV mission will result in a missed opportunity to exercise integrated support of a remote unit assault and waste valuable / limited ISR and CAS assets.

As of August 12, 2016

- Mission ID: S3
- Mission Location: Undisclosed Middle Eastern country
- Mission Brief:
  - Description: Your Unit is tasked to conduct an operational autonomous resupply mission of a supported unit. The cargo consists of POL (petroleum, oil & lubricants), food, and water. This is an operational mission from Camp A to Camp B, which are Coalition "controlled" areas in a Middle Eastern country. You should expect to proceed through six waypoints at a speed of ~2 m/s with mission duration of ~5:00 minutes.
  - Threat Environment: There have been reports of sporadic adversary activity in the area of operation over the last month. This group espouses malicious intent towards Coalition forces, and it is suspected to possess cyber capabilities which could potentially pose a threat to autonomous operations.
  - Likelihood of Mission Success: Despite these reports, only 1 of the last 15 autonomous resupply missions in this region experienced system anomalies. Due the relatively low sporadic nature of these potential adversarial events, autonomous missions should proceed with an expected certainty of success.
  - Risk: This is an operational resupply mission, and the supported unit can persist for a week on its current inventory. UGV safety is always a concern due to the sensitive nature of its components; therefore, a recovery plan exists should it be compromised. The estimated total response time from a Coalition Post to any point along the route is approximately 30 minutes. An attempt will be made to divert an ISR platform to provide over-watch of the UGV recovery.

- Mission ID: S4
- Mission Location: Undisclosed Middle East Country
- Mission Brief:
  - Description: Your Unit is tasked to conduct an operational autonomous (UGV) resupply mission of a supported unit near a Coalition "controlled" border region in a Middle Eastern country. The cargo consists of POL (petroleum, oil, lubricants), food, and water. You will conduct an operational box (return) route resupply mission. You should expect to proceed through six waypoints at a speed of ~2 m/s with mission duration of ~4:00 minutes. Waypoint marker 3 is the resupply location, and the UGV should pause for ~15 seconds at the location to simulate offloading of supplies.
  - Threat Environment: There have been reports of adversary activity in the area of operation over the last month. This adversary group espouses malicious intent towards Coalition forces, and it is suspected to possess cyber capabilities which could potentially pose a threat to autonomous operations.
  - Likelihood of Mission Success: Although no cyber-attacks were officially reported, 10 of the last 20 autonomous missions in this region experienced system anomalies. Due the frequency of these events and potential for adversary activity, autonomous mission success in this area is uncertain.
  - Risk: This is an operational resupply mission, and the supported unit can persist for a week on its current inventory. UGV safety is always a concern due to the sensitive nature of its components; therefore, a recovery plan exists should it be compromised. The estimated total response time from a Coalition Post to any point along the route is approximately 30 minutes. An attempt will be made to divert an ISR platform to provide over-watch of the UGV recovery.

As of August 12, 2016

- Mission ID: S5
- Mission Location: United States, near Nellis AFB, Nevada
- Mission Brief:
  - Description: Your unit is tasked to conduct a joint mission at Nellis AFB, Nevada with the Department of Energy (DOE) for the transport of nuclear materials stored at the Nellis AFB munitions bunkers. You will conduct a box (return) training mission for transport of nuclear material, and you should expect to proceed through five waypoints at a speed of ~2 m/s with mission duration of ~5:00 minutes.
  - Threat Environment: Missions in this area have been protested due to its proximity to the Wildhorse Management Area and Area 51 (see map). Additionally, storage and transport of nuclear materials have been a persistent source of contention with the surrounding communities. However, none of the intel reports suggests activist actions during the mission.
  - Likelihood of Mission Success: Only 1 of the last 15 autonomous missions in the area reported interference with autonomous operations. The reported issue was believed to have been caused by interference from electronic combat range equipment at the nearby Nellis Air Force Range. Due to the low likelihood of activist actions and sporadic nature of past interferences, autonomous mission should continue to proceed with an expected certainty for success.
  - Risk: The mission consists of transport of live nuclear materials and adherence to all DOE protocols. Handling of the nuclear material is a critical safety issue. Once in the transport container, the nuclear material is protected through a series of safety features built into the container for both sensor and remote initiation. Remote initiation with recovery is the normal backup plan. In the event the mission or UGV is compromised, the UGV operator and / or DOE mission commander can remotely initiate cargo container safety features, which will neutralize the nuclear materials making them inaccessible - a very costly decision. A manned crew will then be required to recover the UVG and cargo for proper storage and disposal.

As of August 12, 2016

- Mission ID: S6
- Mission Location: United States, near Nellis AFB, Nevada
- Mission Brief:
  - Description: Your unit is tasked to conduct a joint operational mission with the Department of Energy (DOE) for the transport of nuclear materials stored at the Nellis AFB munitions bunkers to the Nevada Test Site over both public and restricted access roads for live tests. You will conduct a four waypoint out operational mission for transport of nuclear material, and you should expect to proceed through four waypoints at a speed of ~2 m/s with mission duration of ~4:30 minutes.
  - Threat Environment: Missions in this area have been protested due to its proximity to the Wildhorse Management Area and Area 51 (see map). Additionally, storage and transport of nuclear materials have been a persistent source of contention with the surrounding communities. Intel reports suggest the potential for activist groups in the area during the mission timeframe; however, these activists are not believed to possess cyber capabilities which could interfere with the mission.
  - Likelihood of Mission Success: Historically, 10 of the last 20 autonomous missions in the area reported interference with autonomous operations. The issues were believed to have been caused by interference from electronic combat range equipment at the nearby Nellis Air Force Range. Given the frequency of interferences and possible activist activities, autonomous mission success in this area is uncertain.
- Risk: The operational mission consists of transport of live nuclear materials and adherence to all DOE protocols. Handling of the nuclear material is a critical safety issue. Once in the transport container, the nuclear material is protected through a series of safety features built into the container for both sensor and remote initiation. Remote initiation with recovery is the normal backup plan. In the event the mission or UGV is compromised, the UGV operator and / or DOE mission commander can remotely initiate cargo container safety features, which will neutralize the nuclear materials making them inaccessible…a very costly decision. A manned crew will then be required to recover the UVG and cargo for proper storage and disposal.

As of August 12, 2016

Mission ID: S7

- Mission Location: Undisclosed Middle Eastern country
- Mission Brief:
  - Description: Your Unit is tasked to conduct an operational autonomous resupply mission of a supported unit near an ISIS controlled area of an undisclosed Middle Eastern country. The cargo consists of ammunition, food, and water. This is a box (return) mission from / to a coalition camp. You should expect to proceed through four waypoints at a speed of ~2 m/s with mission duration of ~4:00 minutes. Waypoint marker 2 is the resupply location, and the UGV should pause ~15 seconds at the location to simulated offloading of supplies.
  - Threat Environment: Adversaries in the surrounding ISIS controlled area of operation are suspected to possess cyber capabilities, which could potentially interfere with autonomous operations. Additionally, their anti-coalition rhetoric gives reason for concern.
  - Likelihood of Mission Success: Only 1 of the last 15 autonomous resupply missions in this region experienced system issues, and the effected unit completed its mission. The issues were believed to be hardware/software related, and they were not attributed to adversary actions. Due the sporadic nature of these events and potential for adversary activity, autonomous missions should proceed with an expected certainty of success.
  - Risk: This is a mission essential resupply task, and the supported unit needs the cargo to execute its planned objectives. UGV safety is always a concern due to the sensitive nature of its components. If compromised, the UGV would give the adversary significant insights into our tactics and techniques; therefore, a recovery plan exists. If the UGV mission is jeopardized, a manned response team can be mobilized from Coalition Posts to retrieve it with a response time of ~30 minutes. An attempt will be made to divert an armed ISR platform to locate the UGV and provide overwatch and close air support for its retrieval. Failure of the mission will result in a missed opportunity to potentially capture / kill terrorist members and expose the response team and ISR platform to potentially hostile activity.

As of August 12, 2016

- Mission ID: S8
- Mission Location: Undisclosed Middle Eastern country

- Mission Brief:
  - Description: Your Unit is tasked to support a Special Forces (SF) team in a remote operating location of an undisclosed Middle Eastern country. Your mission is the autonomous delivery of a dual mode (manned / unmanned) ATV sized UGV with supplies from a Coalition border camp to a remote location near an ISIS controlled border region. The UGV and cargo, which consists of specialized equipment, food, and water, are required for transport to and assault on a remote terrorist leader meeting location. You should expect to proceed through five waypoints at a speed of ~2 m/s with mission duration of ~4:30 minutes.
  - Threat Environment: Adversaries in the surrounding ISIS controlled area of operation are suspected to possess cyber capabilities, which could potentially interfere with autonomous operations. Additionally, their anti-coalition rhetoric gives reason for concern.
  - Likelihood of Mission Success: 10 of the last 20 autonomous missions in this region experienced some type of issue. Some issues were believed to be hardware/software related, and not attributed to adversary actions. Due the frequency of these events and potential for adversary activity, autonomous mission success in this area is uncertain.
  - Risk: This is a mission essential task, and the SF team needs the transport and supplies to conduct the assault. UGV safety is always a concern due to the sensitive nature of its components. If compromised, the UGV would give the adversary significant insights into our tactics and techniques; therefore, a recovery plan exists. If the UGV mission is jeopardized, a manned response team can be mobilized from Coalition Posts to retrieve it with a response time of ~45 minutes. The armed ISR platform providing overwatch of the objective target will be diverted to locate the UGV and provide overwatch and close air support for its retrieval. Failure of the mission will result in a missed opportunity to capture / kill key terrorist leaders and expose the response team and ISR platform to potentially hostile activity.

Appendix II – Pre-experiment Surveys

Subject Number _____ Test ID_____ Date_____

Demographic Questionnaire

Instructions: Please complete the following statements about yourself by filling in the "blank" and/or placing a "X" in the "blank".

1) Please indicate your age: _____

2) Please indicate your gender: ___Male ___Female ___Other

3) Please indicate your native / first language: _____

4) Please answer the following statements regarding your previous education:
   a. Indicate the highest level of education previously achieved:
      ___High School ____Some College ____Complete 4 yrs. College ____Other
   b. Indicate your previous education major: _____
   c. Indicate your GPA for the highest level of education achieved: _____

5) If currently attending school, please answer the following statements:
   a. Indicate the level of education you are currently pursuing:
      ___Certificate ____Masters ____Ph.D. ____Other (specify):_____
   b. Indicate your current major: _____
   c. Indicate your current GPA: _____

6) Please answer the following statements regarding your work experiences:
   a. Indicate your primary career field occupation: _____
   b. Indicate number of years experience in primary career field: _____

7) Please indicate your employment category: ____Military ____Civilian
   a. If Military,
      i. Indicate your current rank: _____
      ii. Indicate total years in service: _____
   b. If Civilian,
      i. Indicate employment sub-category: ___Government ___Contractor ___Other
      ii. Indicate total years in sub-category: _____
      iii. If prior military, indicate highest rank achieved: _____
      iv. If prior military, indicate number years in military service: ____

Subject Number _____ Test ID_____ Date_____

Personality Questionnaire

Instructions: Read each of the following sections carefully and circle the number that most accurately reflects your feelings or beliefs about yourself for each of those sections.

1. Consider your intelligence as measured by tests such as the SAT, GRE, IQ tests, etc. Read the following question; then circle that best describes you.

   How well do you think you would score on such types of intelligence tests, as compared to others in the United States?

   1. Top 5%
   2. Between the top 5-20%
   3. Between the top 20-50%
   4. Below average

2. Creativity is defined as the capacity to generate new ideas or see new links between old ideas. Using this definition of creativity, circle the option below that best describes how creative you see yourself.

   1. Not creative
   2. Somewhat creative
   3. Creative
   4. Very creative

3. How much do you make time each week to appreciate the creative arts such as architecture, music, fiction, ect.?

   1. Not much of my time
   2. A little of my time
   3. Some of my time
   4. A substantial amount of my time

Subject Number _____ Test ID_____ Date_____

Personality Questionnaire (as of August 12, 2016)

Purpose: The purpose of this questionnaire is to gain insights into your general inclination toward others and/or the actions of others.

Instructions: Read each of the following statements carefully and circle the number that best describes how much you agree or disagree with each statement using the 7-point scale provide below.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree

1. One should be very cautious with strangers.

1 2 3 4 5 6 7

2. Most experts tell the truth about the limits of their knowledge.

1 2 3 4 5 6 7

3. Most people can be counted on to do what they say they are going to do.

1 2 3 4 5 6 7

4. These days, you must be alert or someone is likely to take advantage of you.

1 2 3 4 5 6 7

5. Most sales people are honest in describing their products.

1 2 3 4 5 6 7

6. Most repair people will not overcharge people who are ignorant of their specialty.

1 2 3 4 5 6 7

7. Most people answer public opinion polls honestly.

1 2 3 4 5 6 7

8. Most adults are competent at their jobs.

1 2 3 4 5 6 7

Subject Number _____ Test ID_____ Date_____

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree

9. Most people can be counted on to do what they say they will do.

1 2 3 4 5 6 7

10. I tend to trust people, even those whom I just met for the first time.

1 2 3 4 5 6 7

11. Unless you remain alert, someone will soon take advantage of you.

1 2 3 4 5 6 7

12. Most people would tell a lie if they could gain by it.

1 2 3 4 5 6 7

13. My typical approach is to be cautious with people until they have demonstrated their trustworthiness.

1 2 3 4 5 6 7

14. I usually give acquaintances the benefit of the doubt if they do something that seems selfish.

1 2 3 4 5 6 7

15. Most people pretend to be more honest than they really are.

1 2 3 4 5 6 7

16. I believe that most people are generally trustworthy.

1 2 3 4 5 6 7

Subject Number _____ Test ID_____ Date_____

Personality Questionnaire (as of August 12, 2016)

Purpose: The purpose of this questionnaire is to gain an understanding of your need and/or desire for mental stimulation and creative / complex thinking.

Instructions: Read each of the following statements carefully and indicate whether or not the statement is characteristic of what you believe to be true about yourself. For example, if the statement is extremely untrue of you or of what you believe about yourself (not at all like you), please circle "1" below the question. If the statement is extremely true of you or of what you believe about yourself (very much like you), please circle "7" below the question. You should use the following 7-point scale as you rate each of the statements below.

1	2	3	4	5	6	7
Extremely Untrue me	Untrue of me	Somewhat Untrue of me	Neutral	Somewhat True of me	True of me	Extremely True of me

1. I prefer complex to simple problems.

1 2 3 4 5 6 7

2. I like to have the responsibility of handling a situation that requires a lot of thinking.

1 2 3 4 5 6 7

3. Thinking is not my idea of fun.

1 2 3 4 5 6 7

4. I would rather do something that requires little thought than something that is sure to challenge my thinking ability.

1 2 3 4 5 6 7

5. I try to anticipate and avoid situations where there is a likely chance I will have to think in depth about something.

1 2 3 4 5 6 7

6. I find satisfaction in deliberating hard and for long hours.

1 2 3 4 5 6 7

7. I only think as hard as I have to.

1 2 3 4 5 6 7

8. I prefer to think about small daily projects to long term ones.

1 2 3 4 5 6 7

9. I like tasks that require little thought once I've learned them.

1 2 3 4 5 6 7

Subject Number _____ Test ID_____ Date_____

Personality Questionnaire (as of August 12, 2016)

1	2	3	4	5	6	7
Extremely Untrue me	Untrue of me	Somewhat Untrue of me	Neutral	Somewhat True of me	True of me	Extremely True of me

10. The idea of relying on thought to make my way to the top appeals to me.

1 2 3 4 5 6 7

11. I really enjoy a task that involves coming up with new solutions to problems.

1 2 3 4 5 6 7

12. Learning new ways to think doesn't excite me very much.

1 2 3 4 5 6 7

13. I prefer my life to be filled with puzzles I must solve.

1 2 3 4 5 6 7

14. The notion of thinking abstractly appeals to me.

1 2 3 4 5 6 7

15. I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.

1 2 3 4 5 6 7

16. I feel relief rather than satisfaction after completing a task that requires a lot of mental effort.

1 2 3 4 5 6 7

17. It's enough for me that something gets the job done; I don't care how or why it works.

1 2 3 4 5 6 7

18. I usually end up deliberating about issues even when they do not affect me personally.

1 2 3 4 5 6 7

Appendix III – Post-experiment Surveys

Subject Number _____ Test ID_____ Date_____

Mission Scenario Questionnaire (as of 16 August 2016)

Purpose: The purpose of this questionnaire is to gain insights into your perception of the mission scenario you just completed.

Instructions: Think about the scenario you just completed. Read the following statements carefully and circle the number that most accurately describes your feelings or beliefs for each statement regarding the scenario you just completed.

1. I perceived the consequence of decisions during the mission to be
    1. Not severe at all
    2. Slightly Severe
    3. Somewhat Severe
    4. Moderately Severe
    5. Severe
    6. Very Severe
    7. Extremely Severe

2. Perception of the consequence of decisions during the mission influenced my decision making
    1. Strongly Disagree
    2. Disagree
    3. Slightly Disagree
    4. Neither Agree or Disagree
    5. Slightly Agree
    6. Agree
    7. Strongly Agree

3. I perceived the level of uncertainty about mission success to be
    1. No Uncertainty (100% chance of success; 0% chance of failure)
    2. Minimal Uncertainty
    3. Slight Uncertainty
    4. Some Uncertainty (75% chance of success; 25% chance of failure)
    5. Moderate Uncertainty
    6. A lot of Uncertainty
    7. Complete Uncertainty (50% chance of success; 50% chance of failure)

4. Perception of the level of uncertainty about mission success influenced my decision making
    1. Strongly Disagree
    2. Disagree
    3. Slightly Disagree
    4. Neither Agree or Disagree
    5. Slightly Agree
    6. Agree
    7. Strongly Agree

Subject Number _____ Test ID_____ Date_____

Mission Scenario Questionnaire (as of 8-12-16)

Purpose: The purpose of this questionnaire is to gain an understanding of your perception of the events that occurred during the scenario you just completed and how they may have influenced your decisions during that scenario.

Instructions: Think about the scenario you just completed. Read each of the following statements carefully and circle the number that best describes how much you agree or disagree with each statement using the 7-point scale provide below.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree

1. I was uncertain whether the mission would be successful or not.

1 2 3 4 5 6 7

2. Throughout the mission, I kept generating possibilities about what could be happening.

1 2 3 4 5 6 7

3. I was confident the mission could not be compromised.

1 2 3 4 5 6 7

4. I was on guard during the mission.

1 2 3 4 5 6 7

5. During the mission, I was uncertain as to what could potentially happen.

1 2 3 4 5 6 7

6. I kept thinking some of the events in the mission were unusual.

1 2 3 4 5 6 7

7. During the mission, I felt there was a potential for me to be deceived.

1 2 3 4 5 6 7

8. I was suspicious of events that occurred during the mission.

1 2 3 4 5 6 7

Subject Number _____ Test ID_____ Date_____

Mission Scenario Questionnaire (as of 8-12-16)

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree

9. I was certain of what was going on during the mission.

1 2 3 4 5 6 7

10. There were many times during the mission I found myself wondering about how to interpret the information available to me.

1 2 3 4 5 6 7

11. I was very concerned about the potential for harmful intentions behind some of the events that occurred during the mission.

1 2 3 4 5 6 7

12. I became increasingly suspicious during the mission.

1 2 3 4 5 6 7

13. Throughout the mission, I kept thinking mission success would not be threatened.

1 2 3 4 5 6 7

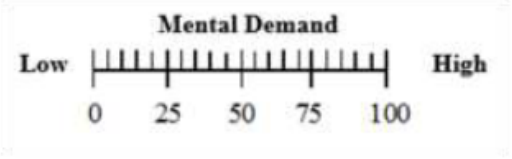
Subject Number _____ Test ID_____ Date_____

Mission Scenario Task Demand Questionnaire

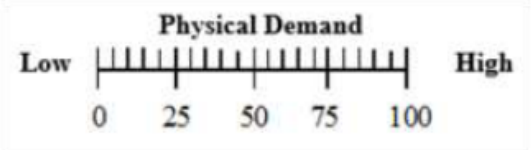
Purpose: The purpose of this questionnaire is to gain an understanding of the demands experienced in performing the tasks in the scenario you just completed.

Instructions: The set of six rating scales below was developed by NASA for use in evaluating experiences in different tasks. Think about the scenario you just completed. Read each question closely and pay attention to the "endpoints" of the associated scale. For each of the questions presented below, place an "X" on the scale that matches your experience with the scenario you just completed, and place a "number" (e.g. 65) in the "blank" corresponding with your placement of the "X". Please note that the "Performance" scale goes from "good" on the left to "poor" on the right.

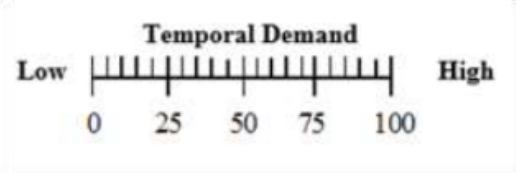
1. How mentally demanding was the task? _____



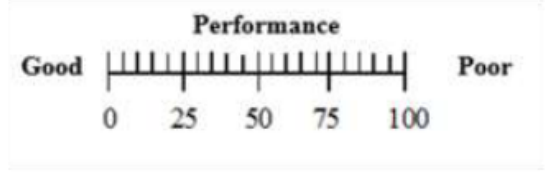
2. How physically demanding was the task? _____



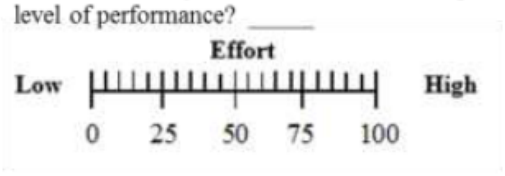
3. How hurried or rushed was the pace of the task? _____



4. NOTE difference in scale: How successful were you in accomplishing what you were tasked to do? _____



5. How hard did you have to work to accomplish your level of performance? _____



6. How insecure, discouraged, irritated, stressed, and annoyed were you? _____