

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188		
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 21-07-2017		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 24-Apr-2013 - 23-Apr-2017	
4. TITLE AND SUBTITLE Final Report: Center For Engineering Transparency: Envisioning Effective Use of Thin-Slicing Through a Model of Human Decision Making With and Trust in Automation			5a. CONTRACT NUMBER W911NF-13-1-0118		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER 206022		
6. AUTHORS Younho Seong, Joseph Nuamah, & Seeung Oh			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES North Carolina A&T State University 1601 East Market Street Greensboro, NC 27411 -0001			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 62815-NS-REP.19		
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT For these objectives stated above, certainly three approaches need to be harmonized to achieve the purpose: 1) how best to assess, filter and/or fuse, and communicate multiple sources of information, and how part/whole information is used in the decision making process, 2) how human trust in automation plays a role in this utilization of such systems and 3) what and how the cognitive processes are induced based on the design of information representation to support commanders' decision making superiority to determine how these should be presented/displayed to be recognized at coup d'oeil.					
15. SUBJECT TERMS human machine interface, neuroergonomics, cognition, information visualization					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Younho Seong
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			19b. TELEPHONE NUMBER 336-285-3734

RPPR
as of 13-Sep-2017

Agency Code:

Proposal Number:

Agreement Number:

Organization:

Address: , ,

Country:

DUNS Number:

EIN:

Date Received:

Report Date:

for Period Beginning and Ending

Title:

Begin Performance Period:

End Performance Period:

Report Term: -

Submitted By:

Email:

Phone:

Distribution Statement: -

STEM Degrees:

STEM Participants:

Major Goals:

Accomplishments:

Training Opportunities:

Results Dissemination:

Plans Next Period:

Honors and Awards:

Protocol Activity Status:

Technology Transfer:

FINAL REPORT

CENTER FOR ENGINEERING TRANSPARENCY: ENVISIONING EFFECTIVE USE OF THIN-SLICING THROUGH A MODEL OF HUMAN DECISION MAKING WITH AND TRUST IN AUTOMATION

AWARD #: W911NF-13-1-0118



Younho Seong, Ph.D.

**1601 E. Market St.
Department of Industrial & Systems Engineering
North Carolina A&T State University
Greensboro, NC 27411**

Sponsored by

US RDECOM ARMY RESEARCH LABORATORY

Department of Defense

JULY 2017

REPORT NO.: ISE 17-YS-01

Performance period: April 2013 – April 2017.

List of students supported, who were awarded (will be awarded soon) through this grant

Ph.D.: Joseph K. Nuamah (expected December 2017), Seeung Oh (Expected 2017)

MS: Gwang M. Kim, Lavis Langley, Marcia Nealy (expected August 2017)

UG students: 2 – Jamir Hart, & Damonte Brimmer

List of publications supported by this grant

Journal articles

1. Nuamah, J., & Seong, Y. (2017). Support vector machine (SVM) classification of cognitive tasks based on electroencephalography (EEG) engagement index. *Brain-Computer Interface*. Taylor & Francis.
<http://www.tandfonline.com/doi/full/10.1080/2326263X.2017.1338012>
2. Nuamah, J., & Seong, Y. (2016 submitted). Evaluation of mental workload and engagement of intuition- and analysis-inducing tasks using continuous wavelet transform of electroencephalography. *A paper submitted to the International Journal of Industrial Ergonomics*.
3. Nuamah, J., & Seong, Y. (2016 submitted). SVM classification of psychophysiological evaluation of mental workload and engagement of intuition- and analysis-inducing tasks. *A manuscript submitted to IEEE Transaction on Man-Machine Systems*.
4. Oh, S., & Seong, Y. (submitted). *Neurological measure of human trust in autonomous systems. International Journal of Industrial Ergonomics*.
5. Oh, S., & Seong, Y. (in preparation). Topographical approach to human trust in autonomous systems. *International Journal of Industrial Ergonomics*.

Conferences and poster presentation.

6. Oh, S., Seong, Y., & Yi, S (2017). Preliminary study on neurological measure of human trust in autonomous systems. *A paper published at the Industrial & Systems Engineering Annual Expo & Conference*. Pittsburg, PA. Institute of Industrial & Systems Engineering Society.
7. Nuamah, J., & Seong, Y. (2017). Neural correspondence to human cognition from analysis to intuition – Implications for display design for cognition. *A paper published at the Human Factors & Ergonomics Society Annual Meeting*. Austin, TX. Human Factors & Ergonomics Society.
8. Nuamah, J., & Seong, Y. (2017). Human machine interface in the internet of things (IoT). *A paper published in IEEE Systems, Man, & Cybernetics 12th*

- Annual System of Systems Engineering Conference*. IEEE SMC Society. Waikoloa, HI.
9. Nuamah, J., & Seong, Y. (2017). Electroencephalography (EEG) classification of cognitive tasks based on task engagement index. *A paper published at the IEEE SMC Conference on Cognitive and Computational Aspects of Situation Managements*. Savannah, GA. IEEE Communication Society.
 10. Oh, S., & Seong, Y. (2017 submitted). Preliminary study on neural correspondence on human trust. *A paper submitted to the Human Factors & Ergonomics Society Annual Meeting*. Austin, TX. Human Factors & Ergonomics Society.
 11. Numanh, J., & Seong, Y. (2016). EEG classification of cognitive tasks based on task engagement index. *A paper presented at the ACM CHI conference Human Factors in Computing Systems*. Denver, CO. ACM SIGCHI.
 12. Nuamah, J., & Seong, Y. (2016). A neurological approach to cognitive state determination. A paper published at *the Industrial & Systems Engineering Annual Expo & Conference*. Institute of Industrial & Systems Engineering Society.
 13. Nuamah, J., Oh, S., & Seong, Y. (2015). Measuring trust in automation: A new approach. *A paper published at 26th Modern Artificial Intelligence and Cognitive Science (MAICS)*. Greensboro, NC.

Poster presentations

1. Nuamah, J., & Seong, Y. (2017). BCI & EEG toward more direct correspondence in human machine interaction. A poster presentation at 2nd Annual meeting of Testing, Evaluation, and Control of Heterogeneous Large-scale Systems of Autonomous Vehicles (TECHLAV).
2. Nuamah, J., & Seong, Y. (2017). Classification of cognitive tasks using artificial neural networks and support vector machines. *A presentation made at the National Science Foundation Emerging Researchers National Conference in Science, Technology, Engineering, & Mathematics*. National Science Foundation.
3. Nuamah, J., & Seong, Y. (2017). Cognitive assessment model for perspective task management. *A poster presented at the 6th Annual College of Engineering Graduate Research Poster Competition*. North Carolina A&T State University.
4. Nealy, M. & Seong, Y. (2016). Ecological approach to model human judgment performance under uncertain environment with autonomous aids. A presentation made at the *1st Annual meeting of Testing, Evaluation, and Control of Heterogeneous Large-scale Systems of Autonomous Vehicles* sponsored by Air Force Research Laboratory.
5. Nuamah, J., Seong, Y., & Stephens, J. (2016). Fuzzy logic-based autonomous decision aids to induce intuition and analysis cognition. A poster presented at the *1st Annual meeting of Testing, Evaluation, and Control of Heterogeneous Large-scale Systems of Autonomous Vehicles* sponsored by Air Force Research Laboratory.
6. Nuamah, J., Seong, Y., Stephens, J., & Homaifar, A. (2016). Task engagement index classification of cognitive tasks using artificial neural network. A poster presented at the *1st Annual meeting of Testing, Evaluation, and Control of*

- Heterogeneous Large-scale Systems of Autonomous Vehicles* sponsored by Air Force Research Laboratory.
7. Nuamah, J., & Seong, Y. (2016). A fuzzy decision making system to access water quality. *A poster presented at the 5th Annual College of Engineering Graduate Research Poster Competition*. North Carolina A&T State University.
 8. Nuamah, J., & Seong, Y. (2016). Cognitive task engagement evaluation using electroencephalography (EEG). *A poster presented at the 5th Annual College of Engineering Graduate Research Poster Competition*. North Carolina A&T State University.
 9. Oh, S., Nuamah, J., Nealy, M., & Seong, Y. (2016). Investigation on the representation of social construct through EEG. *A poster presented at the 5th Annual College of Engineering Graduate Research Poster Competition*. North Carolina A&T State University.
 10. Nuamah, J., & Seong, Y. (2016). Analysis of mental tasks using electroencephalography: A frequency domain approach. *A poster presented at the 5th Annual College of Engineering Graduate Research Poster Competition*. North Carolina A&T State University.

Invited or Featured Presentations

1. Seong, Y., (2016). Correspondence between brain waves and human trust through EEG. Featured speaker at the *1st Annual meeting of Testing, Evaluation, and Control of Heterogeneous Large-scale Systems of Autonomous Vehicles* sponsored by Air Force Research Laboratory.
2. Nuamah, J., Oh, S., & Seong, Y. (2016). EEG-based network classification of intuition and analysis cognition. *A presentation made at the Department of Defense Human Factors Engineering Technical Advisor Group (TAG) 70 Meeting*. NASA Langley Research Center. Hampton, VA.
3. Nuamah, J., & Seong, Y. (2016). CWT & EEG. *An invited presentation made at the Autonomous Control & Information Center, North Carolina A&T State University*.
4. Seong, Y. (2016). Human judgment and decision making processes – engineering transparency? *An invited presentation made at the Autonomous Control & Information Center, North Carolina A&T State University*.

STUDY 1

HUMAN JUDGMENT, TRUST AND ITS' CALIBRATION IN HUMAN-ROBOT INTERACTION – DEVELOPMENT OF SIMULATION TO INVESTIGATE HUMAN AUTOMATION INTERACTION

TABLE OF CONTENTS

1.0	INTRODUCTION.....	10
2.0	LITERATURE REVIEW.....	13
2.1	OVERVIEW	13
2.2	BACKGROUND LITERATURE	13
2.2.1	<i>Human Centered Automation</i>	<i>13</i>
2.2.1.1	CONCEPTS.....	13
2.2.1.2	FRAMEWORKS FOR HUMAN AUTOMATION SYSTEMS.....	15
2.2.2	<i>Research on Automated Decision Aids.....</i>	<i>23</i>
2.2.3	<i>Research on Trust in Automated Systems.....</i>	<i>27</i>
2.2.4	<i>Summary of and Discussions on Background Research.....</i>	<i>38</i>
2.3	General Model of Automated Decision Aids, Trust, and Decision-Making Performance	40
2.4	General Research Questions.....	43
2.5	Summary.....	44
3.0	MODELING FRAMEWORK	46
3.1	Modeling Literature Overview	46
3.2	Lens Model Framework	46
3.2.1	<i>Comparing Systems: The Lens Model Perspective.....</i>	<i>50</i>
3.3	Lens Model of Human Operators with Automated Decision Aids	53
3.4	Details on the Lens Models	56
3.4.1	<i>Lens Model of Human Operator's Judgment.....</i>	<i>56</i>
3.4.2	<i>Lens Model of Automated Decision Aid Performance.....</i>	<i>58</i>
3.4.3	<i>Lens Model of Human Operator's Performance with Automated Decision Aid</i>	<i>59</i>
3.5	Discussions on Models.....	61
3.6	Application of the Modeling Framework to the Study of Trust.....	62
3.6.1	<i>Measuring & Manipulating Reliability, Validity, & Understandability.....</i>	<i>62</i>
3.6.2	<i>Summary of the Three Factors.....</i>	<i>64</i>
3.6.3	<i>The Relationship between Human & Decision Aid Competence & Human Trust</i>	<i>65</i>
4.0	METHODS	69
4.1	INTRODUCTION.....	69
4.2	INDEPENDENT VARIABLES	69
4.3	DEVELOPMENT OF EXPERIMENTAL TESTBED.....	69
5.0	REFERENCES.....	75

LIST OF FIGURES

FIGURE 1. TYPES OF AND LEVELS OF AUTOMATION (PARASURAMAN & SHERIDAN, 2000).	21
FIGURE 2. LEVELS OF AUTOMATION THAT PRODUCED SUPERIOR OVERALL PERFORMANCE IN ENDSLEY AND KABER (1999) SUPERIMPOSED ON PARASURAMAN AND SHERIDAN (2000)'S LEVELS OF AUTOMATION.	22
FIGURE 3. WEB OF FACTORS AFFECTING OPERATORS' USE OF AUTOMATION (RILEY, 1996).	35
FIGURE 4. THREE COMPONENTS IN THIS STUDY.	41
FIGURE 5. FLOW OF INFORMATION AMONG THE THREE COMPONENTS IN THE SYSTEM AND THE ROLE OF HUMAN OPERATORS' TRUST IN AUTOMATED DECISION AID IN JUDGMENT AND DECISION MAKING.	42
FIGURE 6. THE GENERIC LENS MODEL AND DOUBLE SYSTEM DESIGN.	47
FIGURE 7. SINGLE LENS MODEL SYSTEM DESIGN	48
FIGURE 8. N-SYSTEM LENS MODEL DESIGN.	49
FIGURE 9. HIERARCHICAL LENS MODEL WHERE, K = NUMBER OF CUES IN TOTAL, J = NUMBER OF FIRST ORDER JUDGMENT WHICH IS THE SAME AS THE NUMBER OF SECOND ORDER CUES.	50
FIGURE 10. THE LENS MODEL WITH THE STATISTICAL PARAMETERS SUPERIMPOSED.	51
FIGURE 11. APPLICATION OF THE LENS MODEL. ENV: ENVIRONMENT, HO: HUMAN OPERATOR, ADA: AUTOMATED DECISION AID.	54
FIGURE 12. CONNECTED THREE LENS MODELS OF AUTOMATED DECISION AID, HUMAN OPERATORS' JUDGMENT WITH OR WITHOUT THE AUTOMATED DECISION AID SHOWING THE CORRESPONDENCE AND THE EFFECT OF ADA ON THE JUDGMENTS.	55
FIGURE 13. LENS MODEL OF HUMAN OPERATORS' JUDGMENT WITHOUT AUTOMATED DECISION AID.	57
FIGURE 14. LENS MODEL OF AUTOMATED DECISION AID.	58
FIGURE 15. LENS MODEL OF HUMAN OPERATOR'S JUDGMENT WITH AUTOMATED DECISION AID.	60
FIGURE 16. HYBRID LENS MODEL OF HUMAN OPERATORS' JUDGMENT WITH AUTOMATED DECISION AIDS.	61
FIGURE 17. LENS MODEL WITH ITS PARAMETERS CATEGORIZED BY THE TYPES OF COGNITIVE FEEDBACK.	63
FIGURE 18. REPRESENTATION OF THE THREE INDEPENDENT VARIABLES WITH RESPECT TO THE LENS MODEL.	65
FIGURE 19. A TEMPLATE GRID FOR ANALYZING OPERATORS' JUDGMENT PERFORMANCE RELATIVE TO THE AUTOMATED DECISION AID AND THE ENVIRONMENTAL STATE, AND NORMATIVE LEVEL OF TRUST.	66
FIGURE 20. REGIONS OF FALSE AGREEMENT AND FALSE DISAGREEMENT WHEN COMPARING TWO JUDGMENT AGENTS (ADAPTED FROM COOKSEY, 1996).	68
FIGURE 21. THE INITIAL SCREEN OF THE EXPERIMENTAL TESTBED BUILT IN THIS STUDY.	70
FIGURE 22. THE SECOND SCREEN ASKS THE OPERATOR'S DECISION (FIRST DECISION) BASED ON THE DETECTED CHEMICALS DISPLAYED ON THE UPPER RIGHT CORNER.	70

FIGURE 23. THE THIRD SCREEN ASKS THE OPERATOR'S SECOND DECISION AFTER BEING INFORMED FROM THE ADA WITH THE PROBABILITY OF LANDMINE IN THE GRID.	71
FIGURE 24. THE FOURTH SCREEN ASKS THE OPERATOR'S FINAL DECISION.	71
FIGURE 25. THE FIFTH SCREEN ASKS THE OPERATOR'S LEVEL OF SELF-CONFIDENCE ON THE FINAL DECISION MADE.	72
FIGURE 26. THE NEXT SCREEN SHOWS THE POSSIBLE GRIDS THAT THE OPERATOR CAN MOVE ONCE THE OPERATOR MADE THE FINAL DECISION ON THE PREVIOUS GRID (GRAYED GRIDS).....	73
FIGURE 27. THE SCREEN SHOWING THE PATH THAT THE OPERATOR WENT THROUGH. COLORS INDICATE THE LEVEL OF PROBABILITY THAT THE ADA MADE ASSESSMENTS FOR THE GRID. THIS FIGURE IS FOR DEMONSTRATION ONLY.	73
FIGURE 28. THE TRUST QUESTIONNAIRE SCREEN. THE OPERATOR IS ASKED TO FILL OUT THE TWELVE QUESTIONS.....	74
FIGURE 29. THE FINAL SCREEN ASKING TO RATE THE IMPORTANCE OF THE CHARACTERISTICS TO PRODUCE THE FINAL LEVEL OF TRUST.	74

LIST OF TABLES

TABLE 1. THE FITTS LIST (FITTS, 1951).	15
TABLE 2. SHERIDAN’S (1988) LEVELS OF AUTOMATION OF DECISION AND ACTION SELECTION. LEVELS ASSOCIATED WITH AUTOMATED DECISION AIDS ARE ANNOTATED.	16
TABLE 3. LEVEL OF AUTOMATION ADAPTED FROM (ENDSLEY & KABER, 1999).	17
TABLE 4. COMPARISON OF THE LEVELS OF AUTOMATION BETWEEN SHERIDAN'S (1988) AND ENDSLEY & KABER (1999).	19
TABLE 5. MUIR'S FRAMEWORK OF TRUST BY CROSSING BARBER'S TAXONOMY OF TRUST AND REMPEL ET AL.'S MODEL OF THE DYNAMICS OF TRUST. FROM MUIR (1994).	29
TABLE 6. DESCRIPTION OF THE LENS MODEL PARAMETERS OF HUMAN OPERATORS' JUDGMENT WITHOUT AUTOMATED DECISION AID. LMP: LENS MODEL PARAMETERS	57
TABLE 7. LENS MODEL PARAMETERS FOR THE AUTOMATED DECISION AID. * SEE TEXT FOR DETAIL.	59
TABLE 8. LENS MODEL PARAMETERS FOR HUMAN OPERATORS' JUDGMENT WITH AUTOMATED DECISION AID.	61
TABLE 9. OPERATOR'S NORMATIVE LEVEL OF TRUST IN THE AUTOMATED DECISION AID ACCORDING TO THE QUADRANTS IN THE JUDGMENT AND TRUST GRID. O: CORRECT, X: INCORRECT JUDGMENT.	67

1.0 INTRODUCTION

To support human judgment and decision-making performance in dynamic, uncertain environments, many technological advances have been implemented. A technology that has been widely applied in the real world is automation. Automation to assist decision making has been developed in various forms, ranging from simple mechanical devices that display the status of a subsystem in order to assist human in diagnosing the entire system, to very complex forms of automation that may develop their own decision-making strategy, such as using the artificial intelligence techniques. In a domain where the environmental states are not directly observable, various forms of intermediaries have been designed to acquire necessary information about the environment, and to assist human judges in decision-making. It is the job of these automated aids to collect necessary information and process it to generate more useful information that can be considered by human operators. Note that even with a high level of reliability and accuracy in an automated aid, the presence of uncertainty in the environment, which may be transferred into the diagnostic estimates of the automated aids, causes difficulties in judgment tasks. Therefore, it is necessary to design automation that can collect or measure necessary information reliably, and consistently, so that it can positively affect human judgment performance.

In situations where the courses of actions are irreversible or lead to high risk, or the level of uncertainty is very high, the final authority for the judgments tends to be reserved for human operators. Thus, automated systems in these cases are intended to assist human operators by providing informational advice about the states or events of the environment. Then, human operators can either veto or approve the advice provided by the automated systems. That is, human operators, must first estimate the states of the environment, and then have discretion whether to accept or reject the estimates from the automated systems. In this research, we hypothesized that this utilization of automated systems depends in part on the human operators' trust in the systems. As the automated system provides "good" estimates of the environment and demonstrates "good" behavior, one would expect human operators to trust in the automated system more, and consequently rely on the system more.

The study of trust has a long history; however, that history is not rich in empirical studies. Deutsch (1958) made an attempt to define and examine potential characteristics of the term from the sociological perspective. Other studies since then have been performed to understand the role of trust in interpersonal relationships. Recent studies in a sociological context define trust to be a multi-facet concept (Rempel, Holmes, & Zanna, 1985). In parallel to the theoretical contributions from a sociological perspective, research on human operators' trust in automation has been conducted in the human factors community. Based on the definitions from a sociological perspective, two essential studies investigated the role of human trust in continuous process control environment, and producing results generally consistent with those from the sociological study of trust. These studies (Lee & Moray, 1992; Muir & Moray, 1996) examined the role of operators' trust in automated systems. Both studies found that operators' trust was related closely to their utilization of automated systems.

Although the characteristics an impact of human trust has been documented in sociological as well as human factors research (Muir, 1987; Rempel *et al.*, 1985), there has been little research regarding decision-making performance in relation to operators trust in automated decision aiding systems. To address this gap, the research conducted here examined the effect of imperfect automated decision aids on human decision-making performance along with trust in the automated systems. In particular, the main research focus was to manipulate factors which have been previously hypothesized to impact trust, and to investigate the effect of these factors on operators' judgment performance, as well as trust.

Research on people's judgment and decision-making behaviors is very rich in both theoretical and empirical studies. Many decision-making models have been developed to describe people's behaviors from theoretically driven and normative decision-making models like expected utility model to practically driven and descriptive models, such as recognition-primed decision-making model. Among many models which exist today, the Lens Model framework with its extensions (Cooksey, 1996) includes concepts found in classical decision models (how human sample and combine information), as well as describing the relationship between available cues and the true state of the environment. Among the modeling extensions within the Lens Model framework, two system, n-system and hierarchical system model are relevant to the environment of interest in this study: human decision-making supported by automated decision aids. To capture operators' judgment policies, a Hybrid Lens Model was developed to represent operators' judgments, outputs from an automated system, and operators' reliance on the automated output. One important relationship that must be identified in this study to understand the relationship between operators' trust in automated decision aids and utilization of the system is the effect of automated decision aid's estimates on operators' judgment policy. The Hybrid Lens Model developed here is used to provide such information.

Additionally, this research studied a design intervention intended to help calibrate operators' trust in automated systems. Two research streams were considered and linked to identify a way to support operators' judgment and further calibration of trust. Previous research on people's judgment and decision making claimed that providing the task information improved people's judgment performance most. According to Balzer *et al.* (1992), people performed a judgment task when they were provided with the task information including a statistical relationship between the environmental criterion and the informational cues, and a relationship between predicted and actual environmental criterion, than when they were provided with simple performance feedback. From the list of potential characteristics of trust suggested by Sheridan (1988), understandability or transparency were chosen to be useful in both increasing human operators' judgment performance and calibration of trust. Therefore linking these above two, the cognitive information of automated decision aid for the task information was used to accomplish both purposes.

This final report is organized as follows.

First, Chapter 2 reviews literature from three different perspectives on human judgment with, and trust in, automated systems. Description of early studies, i.e., Fitts' (1951) list of function allocation between human and machine or automation are provided to find connections to later studies on automation taxonomies (Parasuraman & Sheridan, 2000; Endsley & Kaber, 1999). Then, trust in interpersonal relationship from sociological perspective is described to form the basis of operators' trust in automated systems from human factors perspective. An emphasis is placed on Sheridan's (1988) description of the definition of trust characteristics, identifying several experimental variables to be controlled. The environment where human operators are supported with automated decision aids is described to discuss several important relationships that must be identified to understand human machine interaction.

In Chapter 3, a detailed description of the Lens Model framework with its extensions (Cooksey, 1996) is provided to model a variety of relationships among the components which are described in the previous chapter. The conceptual relationships between the human operators and the automated decision aid are translated into the parameters within the Lens Model framework. Not only was the separate Lens Model of each component in this study provided, but also a description of Hybrid Lens Model developed is given to understand the relationships. A Hybrid Lens Model describes a model of human operators' judgment with automated decision aids established based on the Lens Model framework. Since the Lens Model framework provides useful methods to capture operators' judgment policy and performance, the important relationships capturing effect of the decision aids on operators' judgment, and reliance on automated decision aids, along with a brief discussion on statistical concepts.

Chapter 4 presents the independent variables to instantiate the effect of various characteristics of automated decision aids on human judgment performance and its relationship with operators' trust in such systems. The characteristics discussed in the previous chapter are experimentally controlled. This chapter also presents the dependent measures that must be collected to evaluate and understand operators' judgment behavior with automated decision aid. Finally, the experimental design and framework are discussed.

2.0 LITERATURE REVIEW

2.1 Overview

Understanding human operators' judgment and decision-making behavior with automated decision aids is important in order to support and increase their judgment performance. There are many possible ways to support human decision making activities, such as designing information displays which are easy to understand, devising training programs, or implementing automated decision aids.

The goals of this chapter are four fold. First is to look at how human operators' judgment and decision-making performance can be affected by imperfect automation technology. Thus, it is necessary to review literature on human centered automation. Specifically, the extent to which human operators are considered in the design of and study of automation technology needs to be reviewed.

Second, there are clear differences between automated systems in supervisory control contexts and automated decision aids which are generally designed to support decision-making activities rather than actually controlling the systems. Thus, review on research on automated systems and automated decision aids will be discussed, specifically from the human centered automation perspective.

Third is the issue of operators' trust in automation. Research on operators' trust in automation is a relatively new area, in which research has been performed primarily in the context of automated systems in supervisory control. A review of operators' trust in automation will be provided, leading to an analysis of judgment performance and trust in automated decision aids.

Fourth, human decision making has been an important issue of study for many years, and researchers have approached the problem of understanding human decision making from a variety of perspectives. Although there exist many decision-making modeling frameworks to date, these models have not been used to represent the situations where human operators are supported with automated decision aids which provide estimates of the environmental states. Therefore, an effort to develop a modeling design based on the Brunswik's Lens Model framework (Brunswik, 1952; Hammond, Stewart, Brehmer, & Steinman, 1975) will be described. A new modeling framework, described in Chapter 3, was developed to fulfill this purpose.

2.2 BACKGROUND LITERATURE

2.2.1 Human Centered Automation

2.2.1.1 Concepts

Automation technology and human operators are both components of complex systems. Many different problems have been documented when humans are not considered at the design stage of automation, such as new forms error (Sarter & Woods, 1995), increased mental load (Wiener, 1988), increased monitoring demand (Parasuraman, Molloy, & Singh, 1993), trust problems (Lee & Moray, 1992) decision

biases (Mosier & Skitka, 1996), skill degradation (Hopkins, 1994), out-of-the-loop performance (Wickens, 1992), low level of situation awareness (Endsley & Kiris, 1995; Wiener & Curry, 1980), cognitive overload (Kirlik, 1993), and masking of incompetence (Hopkins, 1994). Four characteristics can be summarized as important aspects to consider in designing automated systems: its complexity (Sheridan, 1988), its autonomy (Sarter & Woods, 1995), the coupling or interdependence, and its opacity (Woods, 1988).

First, complexity refers to the interaction between and the transitions across a large number of different system elements or modes. It is difficult for human operators to understand systems that are very complex in nature. Such understanding is essential to facilitate in building or maintaining operators' mental models of the systems' behavior.

Second, autonomy or authority refers to automation taking over control of a variety of functions such as decision-making without notifying its decisions to human operators. "Silent" automation indicates the system that provides none or poor feedback about its activities or intentions (Norman, 1990). While autonomous automated systems refer to the systems which can function for long periods of time without further intervention (Woods, 1993), a degree of authority indicates that the systems also is capable of taking over control of the monitored process from another agent if it decides that intervention is warranted based on its perception of the situation and its internal criteria (Sarter & Woods, 1995). Increasing autonomy and authority creates new monitoring and definitions of roles of the components within a system. Human operators as supervisors have to keep track of the status of and changes of modes (Sarter & Woods, 1995).

Additionally, the issue of autonomy brings an issue of responsibility (Mockler, 1989; Moon & Nass, 1998). As Zeide & Liebowitz (1987) noted, the current judicial system indicates that the liability must be traced back to human resource. Diffusion of responsibility is more likely when malfunctions happen. Especially for those systems that human operators are responsible for system effectiveness and safety, human centered automation must remain firmly in command of all components of the system.

Third, the coupling or interdependence refers to the degree of dependence between parts within a system. Coupling is a related property that refers to a system design where a particular input results not only in a particular expected and desired behavior but also has additional 'indirect' side effects (Woods, 1988). The stronger the interdependence or coupling, the more parts are intertwined in terms of individual functions. This creates the problem of latent failures (Reason, 1990), meaning that the negative effect of one part of the system has consequences that are not revealed or activated until some other enabling conditions are met. Not only may this make providing necessary information for diagnosing situations, (i.e., fault diagnosis) difficult, but also this increases the demand on human operators' mental model of automation functioning or connectivity. Without such understanding, humans cannot form clear mental models of how the automated systems should behave, and otherwise they cannot predict the automated systems behavior in the variety of circumstances they may encounter. That is, highly coupled automated systems demand better trained operators who must have better

understanding of automated systems' capability or functioning (Woods, 1996). However, this defeats the original purpose of implementing automation from the beginning.

The above three characteristics lead to the final characteristic: the degree of transparency (Wickens, 1992). High levels of complexity, of very silent automation, and of interdependence eventually make the automated systems of opaque, preventing the operators from observing the inner workings of the systems. Although the display format is a way to resolve this issue of transparency, it would be much better to refer to interface and feedback issues. This term captures the fundamental relationship among the components of a system, the environment and human operators, and context of observation.

2.2.1.2 Frameworks for human automation systems

One method that has been applied to the design of human centered automation is the use of frameworks as taxonomies that address different styles of function allocation between humans and automated systems. For example, Fitts (1951), shown in Table 1, provided a list of disciplines when functions must be allocated between human and machine. Basically, this list is based on the superiority concept in that those tasks that humans do better must be assigned to human operators and vice versa. For example, machines are better at entertaining the same judgment tasks repeatedly based on fixed or certain criteria while human operators are better at making judgments based on inductive reasoning.

Table 1. The Fitts List (Fitts, 1951).

Men (humans) are better at:
• Detecting small amounts of visual, auditory, or chemical energy • Perceiving patterns of light or sound • Improving and using flexible procedures • Storing information for long periods of time, and recalling appropriate parts • Reasoning inductively • Exercising judgment
Machines are better at:
• Responding quickly to control signals • Applying great force smoothly and precisely • Storing information briefly, erasing it completely • Reasoning deductively • Doing many complex operations at once

This list provided a starting point to approach allocating functions between human operators and machines, which further allow the designers of automated systems to make decisions on the types of automated systems. However, later frameworks addressed more explicitly aspects of shared and dynamic function allocation with more complex automated systems.

Sheridan and his colleagues (Sheridan, 1988; Sheridan & Verplank, 1978) proposed a somewhat different kind of list, a scale of degrees to which automated systems are involved with human operators' activities such as generating decision alternatives, decision-making, and action implementation. According to Sheridan's (1992) definition, automation is "the automatically controlled operation of an apparatus, a process, or a system by mechanical or electronic devices that take the place of human organs of observation, decision, and effort." This definition of automation can include a basic form of automation such as simple mechanical device and also more complex form of automation that can generate advanced and processed data for rendering human operator's judgment tasks easier.

Table 2. Sheridan's (1988) levels of automation of decision and action selection. Levels associated with automated decision aids are annotated.

	Level	Description	
High	10	Complete autonomous computer decision and action	
	9	The computer informs the human only if it decides to	
	8	If asked, the computer provides information to the human	
	7	Automatic execution, then necessarily informs the human	
	6	Automatic execution after a certain amount of time	
	5	Execution with the human approval	Automated
	4	The computer suggests one alternative	Decision
	3	The computer narrows down to several alternatives	Aids
Low	2	Human given alternatives of decision and action	
	1	Complete human responsibility in decision and action	

The levels of automation suggested by Sheridan (1988) shows that the number of functions of automated systems gradually increases from a low to a high level of automation and so the role of such systems becomes more critical in performing the tasks. That is, this list shows less human operators involvement as the level of automation increases. The highest level of automation indicates a total supervisory control where the operators may not have the authority to override the automated systems' outcomes or actions. On the other hand, the lowest level of automation indicates the complete responsibility of human operators for all tasks.

Factors such as the human operators' cognitive capability, and the reliability of automated systems, need to be considered in selecting an appropriate level of automation for a situation. For example, there can be cases where adequate performance on critical or important tasks cannot be met with automation technology. In this case, the level of automation must be reduced, and accuracy weighted more heavily than advantages of automation such as speed.

An approach taken by Endsley & Kaber (1999), shown in Table 3, developed taxonomy for levels of automation. The levels of automation are similar to Sheridan's list of automation. In any systems which human operators are assisted with automated systems, there exist four generic functions that must be assigned to either human operators or machines; (1) monitoring – scanning displays to perceive and identify system status; (2) generating – establishing alternatives or options to achieve purposes; (3) selecting – making decisions on a particular strategy; and (4) implementing – implementing the decision into necessary actions. Similar to Sheridan (1988), ten levels of automation were identified varying the degree of involvement of automation in performing tasks, shown in Table 3, and discussed below.

Table 3. Level of automation adapted from (Endsley & Kaber, 1999).

Level of automation		Roles			
		M	G	S	I
1	Manual control	H	H	H	H
2	Action support	H/C	H	H	H/C
3	Batch processing	H/C	H	H	C
4	Shared control	H/C	H/C	H	H/C
5	Decision support	H/C	H/C	H	C
6	Blended decision making	H/C	H/C	H/C	C
7	Rigid system	H/C	C	H	C
8	Automated decision making	H/C	H/C	C	C
9	Supervisory control	H/C	C	C	C
10	Full automation	C	C	C	C

M: monitoring, G: generating, S: selecting I: Implementing. H: Human, C: Computer.

- (1) Manual control: At this level, human performs all the four tasks including monitoring the status of the system, generating alternative, making decisions, and finally implementing the decisions.
- (2) Action support: Computer or machine shares or takes over monitoring and implementing activities, meaning that the system assists the human operator with performance of the chosen actions.
- (3) Batch processing: This is the level of automation that the automated system carries out the human's decision making which is selected among alternatives that are

also generated by the human. Simple automated systems exist such as batch processing in manufacturing operations.

- (4) Shared control: Among the four functions defined previously, selecting an alternative is remained at human operators' full control over. Shared control means that both the system and the human share control over the action implementation.
- (5) Decision support: Decision support indicates that the system support with generating decision alternatives for the humans to select from. However, it must be noted that the human also can generate other or same alternatives which further allow the human to compare it to against those from the automated system. As Endsley & Kaber (1999) noted, this is representative of many expert systems that guide the human through the problem solving process. The difference from the previous level of automation, shared control, lies on the locus of action implementation. At this level, actions are automatically implemented without any human involvement.
- (6) Blended decision making: This is equivalent to the level 5 in the Sheridan (1988) taxonomy of automation in which the automated system only executes the humans decision making only with approval. As is in the previous level, the system executes the decision automatically.
- (7) Rigid system: Rigid system is representative of a system that only a limited number of alternatives are generated by the system. This indicates that the human may or may not have a chance to generate other options.
- (8) Automated decision making: Generation of alternatives and making decisions are entertained by the automated systems as well as implementing the selected option. An intermediate stage is possible where the list of alternatives can be augmented by the human operator. However, the selection process still remains at the automated system.
- (9) Supervisory control: This is the level of automation emphasizing the human operators' activity that monitor and supervise what the automated system is performing. In a highly automated system, the human operator monitors and intervenes the process where necessary.
- (10) Full automation: At this level, the automated system performs every task necessary. Therefore, the human operator is out of the loop and not is allowed to intervene the process.

Crossing the level of automation with the four roles within human-machine interaction, they defined an assignment scheme between humans and automated systems for each level of automation. Allocating these four tasks to either human operators or machines depending on the level of automation, they conducted an experiment to investigate whether different levels or types of automation might have effect on operators' performance. They found that levels of automation which distributed the roles of option generation and implementation between human and/or computer servers had a significant impact on overall system performance. Specifically, the level of automation that combined human generation of options with computer implementation produced superior overall performance during normal operations, as compared to purely manual control and to higher levels of automation involving the computer generation of options.

Reviews of two important studies on the levels of automation are provided. It is apparent that the two taxonomies presented by Parasuraman and Sheridan (2000) and

Endsley and Kaber (1999) are similar in terms of definitions of the roles of the humans and the automated systems. The comparison is shown in Table 4.

Although both taxonomies are very similar and focus on responsibility of decision making task and performing the decisions, Sheridan's (1988) level of automation is more focused on how information flows between the two parts of the system. For example, four levels are related to the presence of decision making notification and the timing of notification. On the other hand, Endsley and Kaber (1999) focused on the assignment the four roles that both agents might take part in controlling.

Parasuraman and Sheridan (2000) devised a simple model for types and levels of human interaction with automation. When designing an automated system, the designer or the decision maker has to make many decisions such as what to automate and how far the system must be automated based on various aspects, such as the operators' mental resources, or a cost and benefit analysis based on the degree of criticality or consequences from which mis-operations or malfunctions may cause. According to Parasuraman and Sheridan (2000), there are several factors to consider when to decide what or how to automate systems, by defining types of or levels of automation, respectively.

Table 4. Comparison of the levels of automation between Sheridan's (1988) and Endsley & Kaber (1999).

Sheridan (1988)	LOA	Endsley and Kaber (1999)
Complete autonomous computers		Full automation
The computers informs the human only if it decides to		Supervisory control
If asked, the computer provides information to the human		Supervisory control
Automatic execution, then necessarily informs the human		Supervisory control, Automated decision making
Automatic execution after a certain amount of time		
Execution with the human approval		Rigid system, Blended DM
The computer suggests one alternative		Blended decision making
The computer narrows down to several alternatives		Decision support
Human given alternatives of decision and action		Batch processing, Action support
Complete human responsibility in decision and action		Manual control

According to Parasuraman and Sheridan (2000), there are four types of tasks relevant for automated systems: Information Acquisition, Information Analysis, Decision Selection, and Action Implementation. They define that information acquisition refers to the sensing and registration of input data. Information analysis involves cognitive functions such as working memory and inferential processes. Examples are the predictor displays (Morphew & Wickens, 1998) for the cockpit and trend displays in process control (Moray, 1997). Decision selection involves varying levels of augmentation or replacement of human selection of decision options with machine decision making while action implementation refers to the actual execution of the action choice.

Among many factors affecting these two questions of types of and levels of automation according to Parasuraman and Sheridan (2000), automation reliability and cost of decision outcomes in terms of risk involved were identified as the most critical factors. For both information acquisition and analysis types of tasks, they suggest a high level of automation if the associated systems are reliable, i.e., detecting signals (system B in Figure 1). This is especially the case of automation where human operators do not have access to the remote sources of information. Therefore, the main factor which impacts the level of automation in these types of tasks is the reliability of the available automated systems. Additionally, feasibility of automation becomes a decisive factor in determining whether these devices should be automated or not. However, there are many cases where it is simply impossible for human operators to acquire and analyze the information, such as in the air traffic control system (ATC) that Parasuraman and Sheridan (2000) used. For example, in the ATC system, it is not feasible for human operators to measure the necessary information such as altitude of an airplane. Additionally, the level of automation for the decision selection type of automation should be low if it involves a high level of risk. In a military target identification environment, where human operators must identify unknown objects based on measured or sensed information, the risk function involved tends to be very high. Therefore, the final decision selection must be remained at the human operators' discretion (system A in Figure 1). Figure 1 shows the Parasuraman & Sheridan's (2000) framework of the levels of and the types of automation.

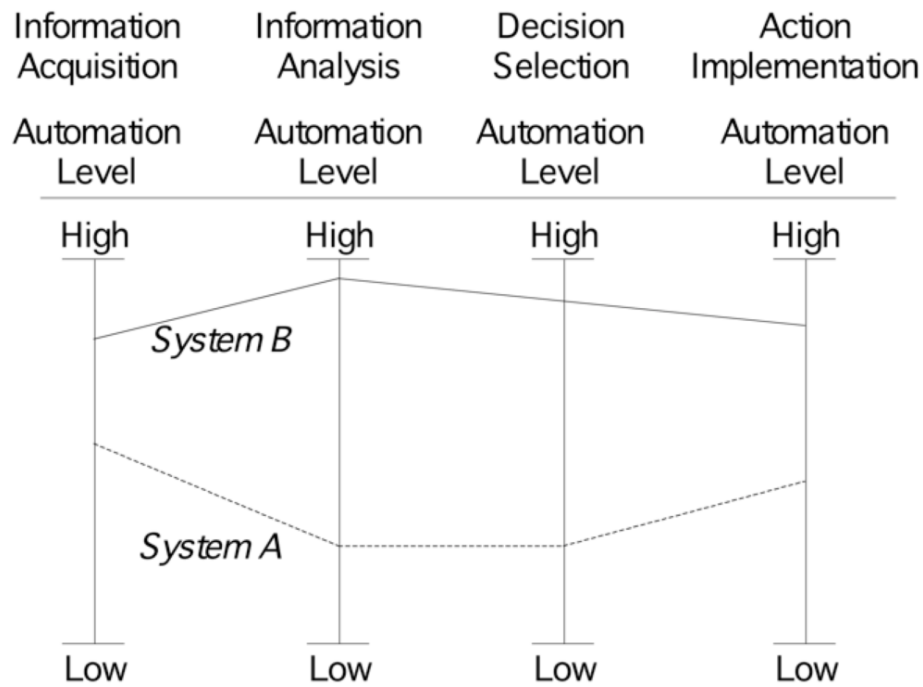


Figure 1. Types of and levels of automation (Parasuraman & Sheridan, 2000).

Based on the literature review, it seems that there is a common ground among the three studies conducted about the types of or levels of automation. Combining Parasuraman and Sheridan's (2000) theoretical study on the levels of automation affected by a variety of factors, with Endsley and Kaber (1999)'s empirical result that operators showed the best performance when operators generate options and machines implement them, results in a recommended set of levels represented on Parasuraman and Sheridan's (2000) taxonomy of levels of automation, shown in Figure 2.

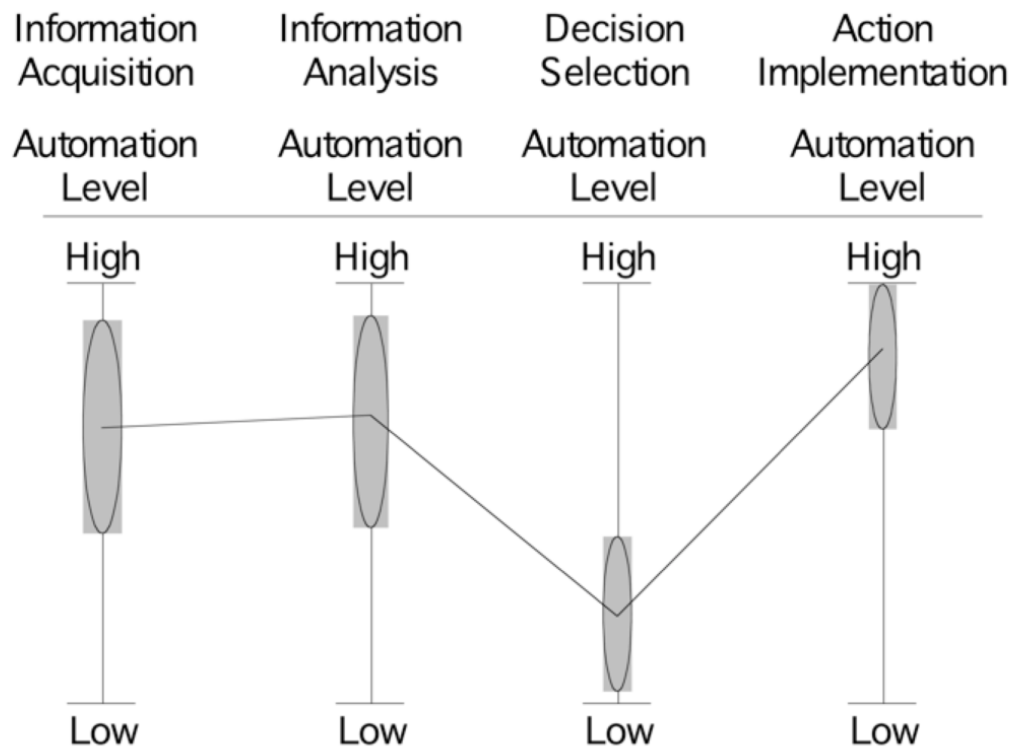


Figure 2. Levels of automation that produced superior overall performance in Endsley and Kaber (1999) superimposed on Parasuraman and Sheridan (2000)'s levels of automation.

2.2.1.3 Summary

Beginning with Fitts (1951) list, the role of machines or computers has been scrutinized to assist better human operators performing interactive tasks with the environment via automated systems. Advancing from the Fitts' philosophy of allocation based on superiority, Sheridan (1988) proposed a level of automation that gradually varies the degree of human operators involvement in performing tasks, starting from manual control in which every task must be performed by human operators, to completely automation-controlled systems in which human operators monitor the automated systems. The main focus of this list lies on the locus of authority for decision-making and execution of the decisions. A similar approach taken by Endsley and Kaber (1999) developed a taxonomy to describe the role of two agents in performing tasks. Their results showed that levels of automation distributing the roles of option generation and implementation between human and/or computer servers had a significant impact on automated system performance.

Additionally, Parasuraman and Sheridan (2000) proposed another approach varying in two dimensions; levels of and types of automation. Simply varying from high to low level, the levels of automation are defined as a continuum. They identified four types of automation, of which could vary in the degree of automation; information acquisition, information analysis, decision selection, and action implementation. The

types of automation can be considered sections of physical components within an entire system. Parasuraman and Sheridan (2000) also described the critical factors such as reliability, or cost of decision outcome, affecting the level of and the type of automation.

In summary, several studies have been discussed to describe approaches to designing automated systems. Whether it is levels of automation or types of automation, it is apparent that the attention of importance of human operators in designing automated systems has been substantially increased.

2.2.2 Research on Automated Decision Aids

One type of automated systems that generate alternatives about the state of the environment or possible courses of actions can be called automated decision aids. Automated decision aids refer to the automated systems that support human operators in problem solving activities or selection of decisions, which require an embedded knowledge base. That is, automated decision aids refer to automation in decision generation and selection according to Parasuraman and Sheridan's (2000) term. In this section, research on automated decision aids will be reviewed. The definitions of automated decision aids, research findings on such systems, and relations of this literature to the automated decision aid in the present work will be discussed.

2.2.2.1 Introduction and definition of automated decision aids

As decision-making environments become more complex and data intensive, the use of automated decision aids is likely to become even more commonplace and more critical to the process. The benefits of automated decision aids and expert systems in terms of increased efficiency for data monitoring and analysis capabilities are fairly obvious. However, the behavioral influences of decision aid use must be assessed to properly assess the benefits, limitations, and costs associated with decision aid development and implementation.

Research has shown that people are very limited in their ability to process information in uncertain environment, especially so with judgment and decision-making (e.g., Tversky & Kahneman, 1974). Researchers from various disciplines have demonstrated that decision aids can assist the human judges making more accurate and consistent judgments (Morrison *et al.*, 1998). Enhanced judgment performance can be attributed to increment in information processing capacity, or to compensating shortcomings of human judgments, such as biases and anchors.

There are many different kinds of automated decision aids that have been designed to support humans in understanding the environment, diagnosing the situations, or engaging in problem solving activities (Rasmussen, Pejtersen, & Goodstein, 1994). For instance, in a system where human operators do not have any access to the actual state of the environment, information regarding the state of the environment is usually collected via various methods such as sensors or other personnel. It is the automated decision aid's job to integrate the information and provide an estimate of the state of the environment. That is, automated decision aids are defined as providing an estimate of the state of the environment based on the information captured from the environment, in

order to support human operators' judgment and decision-making (Morrison, Kelly, Moore, & Hutchins, 1998). A typical example would be a human operator identifying an unknown object as friendly or hostile, based on the information collected via sensors, along with an identification estimate provided by the automated decision aid. Together with the informational cues which are measured or sensed from the state of the environment, the estimate can provide valuable information about the environmental state for the human operator making the final judgment.

2.2.2.2 Research findings on automated decision aids

Although the importance of automated decision aids in assisting human operators' judgment and decision-making tasks has been emphasized over decades, little work has been done empirically in terms of how automated decision aids must be designed to incorporate human operators. The leading force in designing and implementing automated decision aids has been current technology rather than focusing on how people interact with and react to the automated decision aids.

Will (1991) investigated false dependence on technology and performed an experiment with reservoir petroleum engineers in which misleading suggestions were provided about the appropriate methodology to address the given problem situation. The expert information system used by many engineers was to support them in the analysis of pressure buildup problems in backward-chaining by providing both "how" and "why" explanations. Falsified statements in fundamental interpretation of principles in problem solving were provided to investigate their impact on the decision confidence and performance depending on the level of expertise. It is unknown however in what degree the statements from the expert system were falsified from the true solutions. He found no significant difference in decision confidence, and concluded that experts, in general, did not identify errors in the expert system. This result is somewhat consistent with Riley's (1996) in that no significant difference was found between the pilots and college student in activating automation. He also found that novices consistently attributed significantly more to system success than experts, meaning that experts did not view the expert system as helpful in contributing to better job performance or decision-making satisfaction.

One factor that can affect operators' decision-making performance is the complexity of environment. It is obvious that the more complex the environment is, the more difficult the judgment tasks become. Similarly, the complexity of automated decision aids can have impact on operators' understanding of such systems. For example, automated decision aids for the highly structured environment need not to be complex expert systems that require a variety of knowledge built-in (Rasmussen *et al.*, 1994). Simple deterministic type of automated decision aids can perform the task efficiently. Glover *et al.* (1997) investigated whether the decision aid could have an effect on users' acquisition of knowledge and their reliance on a decision aid. They performed an experiment with a computerized determination of tax liabilities, including computation of taxes on net capital gains. The decision aid used in this study, similar to the actual worksheets used by the government, represents a structured task in which participants could be supported by answering initial questions regarding the status of the tax reporter. The decision aid is more or less a decision template which users have to collect necessary

information and enter into the template. Then, the decision aid provided a suggested solution for the users. Therefore, the information collection and analysis tasks remained with the users to figure out what information to look for and how to draw conclusion from the collected information. Their results were consistent with prior research, indicating that the use of decision aid enhanced performance.

On the other hand, highly unstructured problem domain where events are intertwined and multiple causes can create multiple stages of events, need not only expert systems with specific knowledge database, but also human experts. Morrison *et al.* (1998) developed a decision aid to support human operators in understanding environmental situations, making judgments and planning or employing countermeasures in a complex environment. The information provided in this decision support system was that required to make judgments about the environmental state and to evaluate the degree of risk or threat so they could prioritize different activities. For example, the decision support system provided an assessment as to whether or not an unknown object that the human operator intended to identify was a threat along with supporting evidence, counter evidence, and assumptions it was based on. Morrison *et al.* (1998) performed an experiment evaluating the effectiveness of this decision support system and found that the operators reported the system useful for making judgments, and the operators were able to recognize the critical tracks earlier than without the decision support system. Additionally, they found that the operators were able to take more appropriate defensive actions in a timely manner against imminent threats.

In addition to some benefits in terms of performance, it has also been argued that decision aids can facilitate task-related learning (e.g., Libby and Luft, 1993). This argument is based on the claim that structural relationship embedded in the environment can be transferred via the decision aids to the human judges, which further can provide useful information about the environment to establish appropriate knowledge.

When automation is involved in the decision making process, human decision makers may be negatively impacted in terms of the acquisition of knowledge. Glover *et al.* (1997) hypothesized that users with decision aids would acquire less knowledge compared to those without decision aids. The decision aid used by Glover *et al.* (1997) was a simple and highly structured decision aid that provided a suggested solution for a problem in a simple calculation of tax liability. They found that the knowledge score of participants, measured by asking them to write about the domain concept, was significantly less for participants with the decision aid than those without the decision aid. This result was in line with arguments by Wiener (1980), who claimed that one of the negative impacts of automation was operators' skill degradation.

In addition, an important issue to consider regarding automated decision aids is the method to present the outcomes. There can be many ways to simply present information, i.e., graphically, or numerically. In this regard, many studies showed conflicting results on the types of information display on a variety of application areas. Grounds and Wiley (2001) conducted an experiment to investigate the effect of different display interface design on operators' trust in the decision aid. They found no significant

effect of different display types (configural display, bar graph, and alphanumeric) on operators' decision time to agree or disagree with the recommendations. On the other hand, several other researchers have found such an effect (Crocoll & Coury, 1990; Lee, Gore, & Campbell, 1999; Yeh & Wickens, 2001). For example, Yeh and Wickens (2001) performed an experiment focusing on the effect of providing intelligent cuing information to guide attention to targets, and reliability of the automated attention-directing device. They found that the presentation of cuing significantly improve the detection accuracy of low-salience objects. However, the presentation did not help participants in detecting accuracy of expected, high-salience targets. Also, they found that the unexpected target was less likely to be detected when it was presented with a cued object than with an uncued one.

2.2.2.3 Automated Decision Aids in the Current Work

In this research, an automated decision aid typically refers to the types of automation ranging from information acquisition to information analysis according to Parasuraman and Sheridan (2000), which they collectively called information automation. In such systems, data or information collected from the environment is analyzed and fused together to produce further advanced or enhanced information to be useful for human operators making final judgments. Also, it could be categorized as the decision selection type of automation because of the estimates generated by the automated decision aids. For example, the automated decision aid notifies the human operators of the decision outcomes when it decides to do, or upon human operators' request on demand. Note that the human operators are not involved in the decision making process. Because of this reason, the automated decision aids in this study is limited to the information automation in which the decision selection is allocated to the human operators.

Specifically, automated decision aids in the current work refer to systems supporting human operators' decision making activities by integrating sensed information from the environment which the operators do not have any means of access to, and providing sensible estimates of the state of the environments based on a computational algorithm. Therefore, the automated decision aid collect necessary data or information representing the environment, combine them into an estimate in a more comprehensible form of the environmental state. Therefore, automated decision aids support human operators computation and information integration task, which is otherwise a cognitively challenging task. Each set of information is probabilistically related to the environment which means that the information is not completely diagnostic of the environment, and that the information fusion mechanism plays a critical role in generating valid environmental estimates. This is important because it not only affects the quality of the estimates, but also it has an impact on operators' trust in automated decision aids. Consequently, it may further impact on utilization of the automated decision aid.

As Endsley and Kaber (1999) described, level 5 automation, decision support, indicates that the computer generates a list of decision options that the human can select from or the operator may generate his or her own options. This level represents of many expert systems or decision support systems that provide options guidance which the

human operator may use or ignore in performing a task. The automated decision aids in this current work can be categorized as the level 5 in Endsley and Kaber's (1999) taxonomy of the level of automation, and can be any levels between level 2 and level 5 in Sheridan's level of automation. In fact, the level of automation that Endsley and Kaber (1999) described as producing superior overall performance is the same level as the level of automation in this study.

2.2.2.4 Summary

Automated decision aids are defined here as systems which provide an estimate of the state of the environment based on the information captured from the environment, in order to support human operators' judgment and decision-making. Rather than controlling systems, automated decision aids support operators in terms of collecting information, integrating the sets of information into more comprehensible form of alternatives about the environmental states to ultimately support decision making task. The main support is to support human operators' cognitive information integration task.

Previous research on the levels of and types of automation indicated that automated decision aids are high in terms of the level of automation at the information acquisition and analysis, and are low at the decision selection. The automated decision aids in the current work was mapped into two studies on the levels of and the types of automation. The automated decision aid in this current study follows the general definition of automated decision aids, which human operators are still responsible to final decision making task based on the sets of environmental data and the environmental estimate from the automated decision aid.

Finally, several important factors were described that might have influence on design decisions on the level and types of automation. Careful consideration of these factors must be exerted to provide firm support to human operators interacting with uncertain environment.

2.2.3 Research on Trust in Automated Systems

It is needless to say that trust in interpersonal relationship is a key factor in building a good relationship. Such concepts have suggested to many researchers that trust in automated systems may play a key role in the "relationship" between humans and automation. Based on the sociological perspective on people's trust in interpersonal relationship, several studies have been conducted to investigate the role of operators trust in use of automated systems. It is necessary to review the studies on people's trust from the sociological perspective which will pave the way to look at several studies from human factors perspective.

2.2.3.1 Research and Models of Trust from Sociological Viewpoint

Trust is a key component in developing interpersonal relationships. A certain level of trust in others can make interpersonal relationship successful. Trust is more than simple agreement or disagreement between two parties, although accumulation of these interactions can certainly evolve into developing trust or distrust.

There have been several studies investigating the importance of trust in the interpersonal relationship, different aspects of people's trust in interpersonal environment and how they develop their trust in others.

Barber (1983) proposed a definition which includes multidimensional aspects of and necessary attributes of trust in terms of three specific expectations: persistence, technical competence of role performance, and fiduciary obligations and responsibilities. According to Barber, persistence of natural and moral laws provides a foundation of trust by establishing a constancy in the fundamental moral and natural laws. These expectations provide the basic conditions for social and physical interactions. Technically competent performance, on the other hand, supports expectations of future performance based on capabilities, knowledge, and expertise. Barber discussed this dimension of trust referring to the ability of the other partner to produce consistent and desirable performance which can be subdivided to include three types of expertise; everyday routine performance, technical facility, and expert knowledge. These types will be explained later in conjunction with another human factors error nomenclature.

Barber's third dimension of trust, fiduciary responsibility, concerns the expectation that people have moral and social obligations to hold the interests of others above their own. Fiduciary responsibility extends the idea of trust beyond that based on performance to one based on moral obligations and intentions. This dimension becomes important when agents cannot be evaluated because their expertise is not understood, or in unforeseen situations where performance cannot be predicted. Here expectations depend upon an assessment of the intentions and motivations of the partner, rather than past performance or perceived capabilities.

In addition to the dimensions of trust proposed by Barber (1983), Rempel, Homes, and Zanna (1985) emphasized not only components of interpersonal trust, but also the dynamic characteristics of trust toward a partner, regarding trust as a generalized expectation related to the subjective probability an individual assigns to the occurrence of some set of future events. Rempel *et al.* (1985) found people's trust has various aspects that can be developed over time. They defined the term as a multifaceted important intermediate factor in a successful interpersonal relationship. The characteristics of trust they investigated are predictability, dependability, and faith. These are defined both as the characteristics of trust in interpersonal environment and as the developmental stages over time, faith being the final stage which is usually shown in very mature relationship. According to Rempel *et al.*, predictability, which represents the consistency of recurrent behavior and the stability of the social environment, forms the basis of trust early in a relationship. As interpersonal relationships progresses with further experience, dependability, which represents a more common understanding of the stable dispositions, becomes an important basis of trust between humans and focuses on an evaluation of the qualities and characteristics attributed to the partner. In this stage therefore, the centroid swings and shifts away from the evaluation of the partner's dispositional attributes to the partner. Faith, on the other hand, describes the aspects of trust or belief that must go beyond the available evidence to accept a given supposition as truth. Their survey results showed that these characteristics are developed in the order as the relationships mature.

Results from other studies (e.g., Barber, 1983; Rempel *et al.*, 1985) are generally in consensus in that people develop their trust gradually based on the evidences, and once betrayed, take long time to recover.

2.2.3.2 Frameworks of Trust in Human-Machine Systems

Based on the sociological definitions of trust described in the previous section, Muir constructed a model of human trust in automation by incorporating the dimensions of trust proposed by Barber and three more dimensions from Rempel *et al.* (1985). According to Muir's interpretation, all three of Barber's (1983) meanings of trust seem applicable to the human-machine relationship and become a basis for the framework. Thus, Muir's work forms the first "transition" of the concepts of trust in human-human relationships to human-automation relationships. Muir's work in fact suggests that this is an identity-transform, i.e., that the human-human trust concept extends directly to the human-automation relationship case.

Table 5. Muir's framework of trust by crossing Barber's taxonomy of trust and Rempel *et al.*'s model of the dynamics of trust. From Muir (1994).

Expectation	Basis of expectation at different levels of experience		
	Predictability	Dependability	Faith
Persistence			
Natural physical	Events conform to natural laws	Nature is lawful	Natural laws are constant
Natural biological	Human life has survived	Human survival is lawful	Human life will survive
Moral social	Humans and computers act 'decently'	Humans and computers are 'good' and 'decent' by nature	Humans and computers will continue to be 'good' and 'decent' in the future
Technical competence	<i>j</i> 's behavior is predictable	<i>j</i> has a dependable nature	<i>j</i> will continue to be dependable in the future
Fiduciary responsibility	<i>j</i> 's behavior is consistently responsible	<i>j</i> has a responsible nature	<i>j</i> will continue to be responsible in the future

Muir (1994) also identified one of Barber's aspects of trust, technical competent performance, with Rasmussen's (1983) taxonomy of behavior; skill-, rule-, and knowledge-based behavior. Interpreting Rempel *et al.*'s (1985) model as a hierarchical stage model as being able to account for changes in operators' trust as a result of experience on a system, where trust develops over time, Muir produced a framework by crossing Barber's (1983) dimensions of trust, with Rempel *et al.*'s (1985) framework; this crossed relationship among these various dimensions is shown in Table 5. Table 5 shows

an integrated model of trust created by Muir (1994) by crossing Barber's (1983) multidimensional characteristics of trust and Rempel *et al.*'s (1985) model of the dynamics of trust to investigate the meanings of and the role of trust in human-machine relationships.

Based on the definitions of trust in interpersonal environment, Muir and Moray (1987) applied the same definition to a human machine environment where operators had to control multiple system components which can be switched between manual and automatic operation. Each component could be controlled in manual or in automatic in which the human operator's task in automatic mode of control is simply to monitor the automation.

Additionally, Sheridan (1988) provided a list of attributes of trust specifically for human machine environment.

Reliability: This factor refers to a system of reliable, predictable, and consistent functioning. Most of the prior definitions of trust addressed this attribute as the first step in developing trust, based on the premise that a person who behaves in a consistent manner will be trusted easily. Lee and Moray's (1994) experiments showed the importance of reliability affecting the level of trust in automation, by manipulating by the error magnitude. However, Lee and Moray's definition of reliability is broader than Sheridan's. While the latter indicates an automated system's consistent functioning, the former includes the degree to which automatic controlled values are close to the target values as well.

Validity: The validity of an automated system refers to the degree to which it produces correct output. This implies how faithfully the automated system can estimate the environmental state. In operations in which information is integrated from multiple inputs or sensors to produce estimates of the environmental state, validity refers to the distance between the environmental state and the output from the automated system. It seems intuitive that the automated system that produces more valid outputs to the human operators will be trusted more.

Robustness: Robustness supports expectations of future performance based on capabilities and knowledge not strictly associated with specific circumstances that have occurred before. Sheridan (1988) stated this as "meaning demonstrated or promised ability to perform under a variety circumstances." For example, consider the sudden increase of water inflow into a system; this may cause an automatic valve system subject to malfunction. If the automated system was designed to handle this situation whether it is expected or unexpected, it will still provide a useful way to control the situation. This should increase operator's trust in the system.

Glover *et al.* (1997) performed an experiment investigating the effect of inferior decision aid on users' performance in tax calculation domain. As discussed in the previous section, the decision aid used in the study was simple representing highly structured environment. However, the decision aid was designed to mainly focus on the "typical" cases. That is, the decision aid was not robust to consider a variety of cases that

can exist in the real world. Their experimental result showed that participants' performance was significantly better with the decision aid than those without the decision aid, but only for those typical cases. For non-typical cases, participants without the decision aid significantly outperformed those with the decision aid. From the result, it appears that if the decision aid is not robust to handle many different cases, it is of no use in supporting human operators' judgment and decision-making activities.

Transparency: Transparency refers to the degree to the extent which the inner workings or logic of the automated systems are known to human operators, to assist their understanding about the system. Opaque systems become another source of uncertainty that can have negative impact on operators' judgment and decision-making performance. Unless human operators understand how the automated system performs, they may either disregard or totally rely on the automated system without questioning its capability. Therefore, providing a method for the operators to understand the automated systems better can increase the chance they learn to appropriately rely on the automated systems.

Understandability: The operator's understanding of the automation is important for the expectations that make him trust and use the system appropriately. The operator's intervention will be better and more timely when his trust is well calibrated to the actual trustworthiness of the system. The construct of understandability is related to development of an appropriate mental model. In designing a machine to aid a human operator, understandability is affected by the degree of transparency of the system which the operator can "see" the underlying mechanisms through the interface. Hanes and Gebhard (1966) conducted a series of studies specifically directing to the question of man-machine interaction. The task setting was a highly realistic simulation of a Navy Combat Information Center, and their subjects were experienced naval commanders. Data were obtained from officers in a variety of scenarios with common features of a computer aid which had varying degrees of impact on the solution of the problem of defending the simulated ship from simulated air attacks. They found that officers were much more willing to accept the computer recommendation if they agreed with the logic used and/or chose the logic themselves, if they understood the computer program or if the recommendations were unambiguously displayed. From this study along with others, we can see that operators' knowledge about the automated system plays a critical role in determining reliance of decision aiding technology.

Another way to increase the level of operators understanding is to develop an appropriate display interface design. The interface design presumably matches the operators' concept of automated systems, so called mental model. Many studies (e.g., Vicente & Rasmussen, 1992) have claimed that an interface embedded with relationships among the components can enhance operators' understanding, which consequently contribute to increase in performance. On the other hands, Grounds and Wiley (2001) conducted an experiment to investigate the effect of different display interface design on operators' trust in the decision aid. They found no significant effect of different display types (configural display, bar graph, and alphanumeric) on operators' decision time to agree or disagree with the recommendations.

Hanes and Gebhard's (1966) results showed that officers were more willing to accept the computer recommendation if they believed the input data were good quality, or

if the recommendations were unambiguously displayed. It may be that the usefulness of any system depends on the manner in which system information is presented. In the case of a decision aid which produces the estimate of the environment, there must be a suitable presentation of the uncertainty regarding the system's inferences. This method of presentation becomes the window of opportunity for the operators to reduce the additional uncertainty created by having the automated decision aid.

Usefulness/Utility: The usefulness of data or machines means responding in a useful way to create something of value for operators, eventually developing into trust. One study by Morrison *et al.* (1998) examined the usefulness of decision support systems within the tactical decision-making environment. They provided several decision-making support modules that assisted human operators to understand the situations and making judgments. Conducting an experiment with two operators as a team, they found that the operators tended to rate those parts of the decision-making support system modules that supported quick decision making, and thus were more useful, higher than other modules.

This list seems comprehensive in that it includes factors beyond those of Muir (1994), or of Rempel *et al.* (1985), that may have effect on operators' trust indirectly. For example, consider understandability and transparency. These factors may not be the ones that operators directly access when they evaluate the level of trustworthiness of an automated system. However, as they understand the automated system better, they can grasp knowledge about the pros and cons of the system and when to rely on the system. Therefore, it seems that Sheridan's list of trust attributes provide comprehensive understanding of operators' trust and systematic characteristics of trust that may cause operator's trust to develop.

2.2.3.3 Research on Trust in Automated Control Systems

Based on the research framework to study human trust in human-machine interaction setting, several studies performed experiments in a continuous processing domain.

Among the prior research, two studies, Muir and Moray (1994), and Lee and Moray (1992), made attempts to identify the role of operator's trust in automated systems. Further, these two studies showed interesting results on the relationship between the operator's trust in automated systems and their reliance on the system measured by the use of the system. Conventional wisdom implies that the use of or reliance on automation becomes higher as the human operators trust the system more. For example, Muir and Moray (1994) performed a process control simulation experiment and the results supported this perspective, showing a positive regression coefficient between operator's trust and the use of automation. They also concluded that as the level of operator's trust suffers, their use of automated system decreases. Lee and Moray (1992) conducted very similar experiments and showed that the operators tended to use less of the automated system as the level of trust decreases.

First, having defined operators' trust in automation as a simple additive model of three characteristics, Muir and Moray (1996) conducted a series of experiments to investigate the validity of Muir's (1994) definition of operator's trust in automation and

the relationship between the properties of the automation, trust, and human intervention. The main systematic factor, reliability of the automated system, was controlled by adding either constant or variable errors. These errors were generated and added to the operators' input, the actual process, and the automated system's output. Note that they differentiated between the errors originated from controlling activities and the errors in the display. The level of trust was subjectively approached through a questionnaire. They found that human operators were able to express their trust in automation quantitatively and also found that Rempel *et al.*'s three characteristics were representative to estimate operators' general level of trust in automation. In terms of Barber's (1983) taxonomy of trust, Muir and Moray (1996) stated that competence was the better single predictor of trust. Additionally, they found that an "honest" automated system was trusted more than the other two systems which were contained either constant or variable errors. An interesting finding was that the two "poor" systems were not different in terms of operators' trust in the system. In terms of the dynamics of trust development, they found that faith factor played a critical role in early stage of interaction, then predictability followed by dependability. It is quite contradictory to what Rempel *et al.*'s (1985) results demonstrated. Rempel *et al.* (1985) showed that predictability was the key role in an early stage, then dependability followed by faith.

The second experiment by Muir and Moray (1996) was performed to investigate the differential effect of consistent or variable error on operators' trust in automated systems, and the relationship between the level of trust and monitoring activities. They found first that trust was greatly reduced by small control errors, but was increasingly less sensitive to larger control errors. Operators' use of automated system was also affected by the magnitude of the control error. They also found that participants engaged more on monitoring activities when they trusted the automated system less. This is consistent what Sheridan and Hennessy (1984) suggested that operators would spend more time monitoring automated systems they would not trust.

Second, Lee and Moray (1992) conducted several similar experiments to investigate relationship between human operators' trust in automation and their use of the automated systems. Automation failures were manipulated in two dimensions; length and size. They controlled the length of automation, transient vs. chronic, failure to investigate the dynamic characteristic of trust. Also, this would reveal the length of recovery affected by different length of failures. Automation failures were varied in size also to see any differential effect on trust in automated systems. They showed a strong correlation between human operators' trust in automation and their use of the automatic controller, as well as an inverse relationship between operators' trust and their monitoring activities. Participants showed a decreased level of trust in automation when faults occurred in the automated systems and engaged less in monitoring the automated system under the automatic mode. These results seem correct in that as the level of reliability decreases, human operators should trust the automated system less. They found out, first, that human operators were able to maintain their high level of overall performance, despite the effects of a fault in the automatic controller. This is consistent result with Muir and Moray's result. Participants were able to maintain high level of performance while

interacting with an unreliable automated system. In this case, the degree of reliability was defined as a percentage variation from the targeted values.

A behavioral counterpart of operators trust in an automated system is an index of reliance, use of automation. This is based on the assumption that the more operators trust, the more they use. Riley (1996) investigated a variety of factors affecting the human use of automation simultaneously. He performed an experiment using a simple computer game which participants had to control various simple multiple tasks part of which they could relinquish the control to automated system. Automation reliability (90%, and 50%) was controlled by setting the probability that the automation would perform a task correctly. The automation faults were distributed over time and the experimental conditions, which were different in the level of workload and in the level of uncertainty. The level of uncertainty in a task was controlled by introducing abnormalities so that the automated system could not perform the task appropriately. The level of workload was also controlled by making another separate task more complicated. The results showed that participants demonstrated a bias toward manual control, and that both college student and pilots did not delay turning on automation after a failure, and continued to rely on failed automation. Pilots showed greater tendency toward the use of automation under the lower risk. In a subsequent experiment focusing on the dynamics of trust, Riley controlled the level of information that participants were given about the automation prior to performing the experiment. These conditions differed in whether participants were provided with information regarding automation reliability, state uncertainty, or both to reveal the contributions of each element to their automation use decisions. He found that both state uncertainty and trust affect automation use decisions, but only early in the participants' experience with automation. Based on a variety assumed relationships affecting operators' use of automation, he found several interesting relationships, shown in Figure 3.

Additionally, Itoh and Tanaka (2000) developed a mathematical model of trust in automation in which a variety of combination of the three characteristics proposed by Rempel *et al.* (1985) were used the evaluation criteria to determine whether to rely on the automation or not. The major point that makes this study different from others is that this study considered the state of the automated systems in determining trustworthiness, thus addressing the concept of calibration of trust.

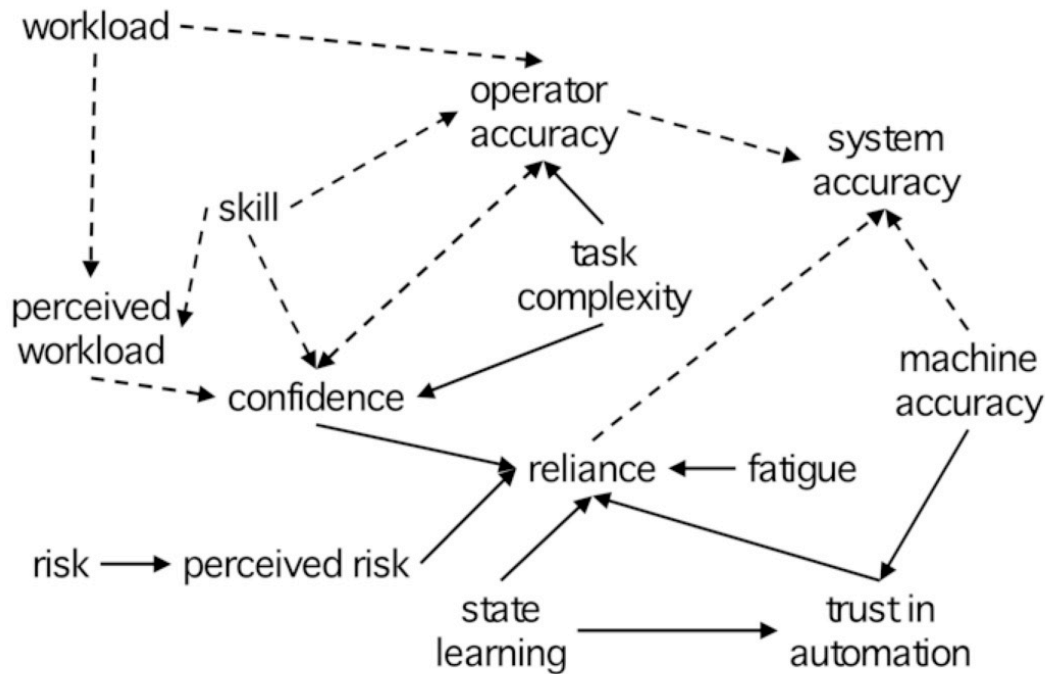


Figure 3. Web of factors affecting operators' use of automation (Riley, 1996).

Although those research projects concentrated on human trust in automation based on the general understanding of supervisory control tasks, the importance of the trust concept seems applicable and has been attempted to define the role of trust in other domains, such as computer supported cooperative work (Jones & Marsh, 1997; Christianson and Harbison, 1997), decision-making in management (Lerch & Prietula, 1989), medical diagnosis expert system (Moffa & Stokes, 1996), computer security problems (Beth, Vorcherding, & Klein, 1994), and also, trust in e-commerce (Gefen, 2000).

2.2.3.4 Research and Models of Trust in Decision Aids or Expert Systems

Specific empirical research on operators' trust in automated decision aids is somewhat more limited than general research on trust in automated systems. Several studies used different types of decision aids that might have some relevance with the issues in this study.

As discussed before, Glover *et al.* (1997) conducted an experiment investigating the relationship between the decision aids and the humans. Among the results, they showed that the users' performance was significantly better with the decision aid than without the decision aid. This was only true for those typical cases. Reverse was the truth for those cases that were not typical. The result means that participants were able to perform the task better without the decision aid for the cases for which the decision aid was not designed and was not expected to encounter. Also, it indicates that participants relied on the decision aid inappropriately for those cases that the decision aid could not

perform. In conclusion, decision aids that are inferior for non-typical cases can cause operators to inappropriately rely on the automation, which consequently results in the low level of performance.

For instance, one study by Dzindolet *et al.* (1999) investigated different aspects of automated decision aids on operators' identification performance and revealed that participants decided to ignore the decisions by the decision aid when no feedback was provided, and when they were told that the decision aid was actually human rather than an automated aid, regardless of the quality of the decision aid (better than or equal to participants performance). However, the percentage of participants who decided to ignore the support from the decision aid when they were told that the decision aid was actually an automated decision aid was not also promising, regardless of the presence of the feedback or the quality of the decisions. The average percentages of participants who decided to ignore the decision aid were 83% and 81% for the feedback and no feedback condition, respectively.

Cohen *et al.* (1998) proposed a qualitative model of trust that allows for the operators' ability to distinguish the situations in which the automation should exhibit better or worse performance. They proposed the Argument-based Probability Trust (APT) model, which borrowed from terminology used in Toulmin's (1958) analysis of logic and argument construction. The main idea of Toulmin's model of argument consists with five components: warrant, grounds, qualified claim, backing, and rebuttals. The final outcome or conclusion can be supported by evidence or grounds that should necessarily be linked by warrant. A warrant refers to a specific kind of links or rules, i.e., if-then rule, that grounds satisfy the claim. Additionally, the existence of backing provides theoretical or empirical explanation of the warrant. Using the model of argument to interpret operators' trust in automation, they described trust in decision aids based on the event trees, which consist with nodes and branches. Branches represent possible alternatives of a factor of the environment that can affect operators' judgment and their trust while nodes represent the user's trust in the decision aid at that point. The level of user's trust is calculated by assigning user's belief on each branch.

Lerch and Prietula (1989) conducted an experiment consisted of traditional financial management decision problems to investigate effect of inferior quality advice on people's judgments. As might be expected, there was a loss of trust in the agent resulting from faults in the advice, and the recovery of trust was slower than that in performance. They called this "inertia", similar to what others have called a "hysteresis loop". They also found that it was more difficult for humans to recover trust after once it was lowered because of the failures than to build trust initially. In Lerch and Prietula (1989), the level of performance measured by the level of confidence in the decisions the participants made deteriorated after the wrong advice, and never returned to the level of performance where it was before the wrong advice, even at the end of trials.

Finally, Bisantz and Seong (2001) defined different types of failures that could occur in information acquisition and information analysis types of automation, acknowledging that systems could fail due to different intents. For instance, system

components or automation may fail due to unintentional hardware or software difficulties. In some environments, failures may be introduced intentionally, through acts of sabotage. They conducted an experiment investigating effects of different types of failure on operators' decision making performance and trust in two informational sources. One informational source, information window, provided sensed information by various sensors while the other, decision-aid window, provided the environmental estimate of an object's identity. Participants were told that failures due to hardware or software failures or sabotage are possible in addition to the third condition where the probability of decision aid fault was not mentioned. Note that operators could select whether to consult with the automated decision aid. Results showed significant impact of fault condition on information window use. Participants in the sabotage condition seemed least likely to select the information window on a track-by-track basis, and participants in the control condition the most likely, while across the three conditions there was a trend toward decreased information window use.

They concluded that there appeared to be a trend for participants in the sabotage condition to make less use of the information window than participants in the other two conditions. Also, participants showed a trend across conditions to make less use of the information window and more use of the decision-aid window, indicating that participants attributed the failure to the information window rather than the decision-aid window.

2.2.3.5 Measurement of Trust

While many studies (e.g., Larzelere and Huston, 1980) have justified the use of measurements using rating scales, it is also obvious that different measurement schemes, (e.g., percentage of time that operators used automatic controllers, and frequency of monitoring activities) should be developed to predict the operators' actions in a human-machine interaction environment. Several studies of trust have been conducted from this perspective.

As discussed in the previous section, while studies on people's trust in an interpersonal relationship are rich in the theoretical perspective, little has been done on empirical measurement of people's trust.

Rempel *et al.* (1985) developed an 18-item questionnaire focusing on three characteristics of trust. An important question is how to validate whether a questionnaire measures a construct such as trust. This be less of an issue for human factors oriented studies of trust however, for which the construct of trust in automated systems can be associated with some qualitative behavioral measures, such as reliance on the automated system.

To date, there is only one study systematically examining the multi-dimensional characteristics of trust and identifying appropriate words or terms to describe people's trust in automated systems. Jian *et al.* (2000) performed an experiment in which participants compared and rated according their subjective relevance to the various characteristics of trust that they defined based on the sociological studies (Rempel *et al.*,

1985). In addition to those characteristics defined in the previous studies, they found several words or phrases that could describe the term trust or lack of trust. As shown in Appendix B, the questionnaire consists with twelve questions to which participants have to respond on Likert-type scales. Five questions are negative describing the automation's behavior that can cause low level of trust while 6 questions describing positive aspects of behavior. This questionnaire has been used in assessing the operators' trust in automated decision aids (Bisantz & Seong, 2001). Bisantz and Seong (2001) found that the questionnaire based on fault condition was sensitive to difference in task conditions, and also that the different levels of fault conditions were affecting participants' subjective responses of trust differently.

Kim and Moon (1998) performed a similar study investigating trustworthiness of cyber-banking system interface, and found that the words that are closely correlated with system's trustworthiness are dependable, reliable, realistic, and uniform.

2.2.3.6 Summary

Research on people's trust in interpersonal relationships reveals that trust is not a simple character that can be easily assessed. Rather, it has multiple facets each of which may require different evidence to grow and which may require substantial period of time to reach at a level where individuals can truly trust the other partner (Rempel *et al.*, 1985). The characteristics identified provided us with a better understanding of the concept of trust in interpersonal relationship which can be extended to human machine interaction perspective.

Literature seemed to suggest that the term within the human automation interaction domain also has multiple characteristics associated with it. As a definitive study, Sheridan's (1988) characteristics of trust in automated systems are described. Empirical research and models of trust in automated control systems are described to provide better understanding of the studies on operators' trust in automated systems. From these studies, it seems that there is common consensus between these studies that the use of automated systems is tied to the level of trust. Finally, existing methods of measuring trust were described.

2.2.4 Summary of and Discussions on Background Research

In the previous sections, three major issues within the human machine interaction paradigm have been discussed to raise important research questions regarding human operators' judgment behavior and trust in conjunction with automated decision aids.

First, a discussion on automated systems is provided with having in mind that different types or levels of automation may stimulate or emphasize different aspects of operators' trust in automated systems, which consequently may have different impact on operators' use of the automated systems. In this perspective, human centered automation was discussed to provide deeper understanding of the types of and levels of functions that human operators interact with the decision aids. Studies regarding the taxonomy (e.g., types and levels) of automation seem to indicate that a variety of combination of different levels and types of automation is possible depending on a variety of factors that can have impact on the decisions on the automation. Among these various alternatives of

automation, automated decision aids are characterized as the systems that every type of automation but the decision selection is designed at a high level of automation. Finally, the automated decision aid in the current research was defined and described to outline differences from the definitions provided in the previous studies on the types of or levels of automation.

Several studies were described to introduce different aspects of automated decision aids that may have impact on the operators' judgment performance or the use of such systems. There seem to be a positive relationship between "imperfect" automated decision aids and operators' performance, which is somewhat consistent to the results from the previous studies on operators' performance with automatic controllers. However, those decision aids used were designed to support human operators problem solving activities. Therefore, it is difficult to identify whether there was any differential effect of the different magnitudes of failures on operators' performance. In other words, we need to first be able to evaluate the competence of the automated decision aids, then it may allow us to understand the differential effect on operators' judgment and decision making performance. One important issue regarding the relationship between operators' judgment behavior and performance has been identified in this process: operators' trust in automated decision aids.

Second, research on trust was described to establish the concept of and the role of trust within this research framework. First, trust in interpersonal relationship was described. Established from the sociological perspective on trust in interpersonal relationship, discussions on human operators' trust in automation together with justifications on the importance of this study on operator's trust in automated decision aids have been provided. Briefly, two important studies (Lee & Moray, 1992; Muir, 1994) revealed that operators' trust in automated control systems plays a critical role in determining the use of automated control systems. Discussions on two streams of studies on trust in automation in general were provided: trust in automatic controllers, and trust in decision aids. Rather than using automated controlling systems, several studies used the decision aids in investigating operators' knowledge acquisition (Glover, Prawitt, & Spilker, 1997), decision confidence (Will, 1991), self-serving or automation bias (Dzindolet, Pierce, Beck, & Dawe, 1999), trust in expert systems in financial domain (Lerch & Prietula, 1989). The main difference between these two automated systems lies on the methods of engagement with the automated systems. This is important because it can be one of the objective measures to interpret operators' trust in such systems. First with the automatic controller, the human operators have to make conscious decisions to engage in automatic mode of control to at least to understand the capability of automation, or how well the automated system performs the tasks compared to what is suppose to do. On the other hand, with automated decision aids, operators do not have to rely on the estimates to understand various characteristics of the automated decision aids. Rather, operators have more chances to evaluate different aspects of the systems, such as its capability, competence, or reliability, without actually engaging it. This can allow us to understand the relationship between the operators' judgment performance, various characteristics of the automated decision aids, and trust in such systems.

Finally, issue of measuring trust in automated systems was discussed to introduce existing methods from both qualitatively and quantitatively.

Having reviewed important issues in human machine interaction, specifically automated systems, let us turn our attention to constructing general model of human operators' judgment and trust in conjunction with automated decision aids.

2.3 General Model of Automated Decision Aids, Trust, and Decision-Making Performance

The automated decision aid in this current study is defined as "Decision-making" level within Endsley and Kaber's (1999) taxonomy of automation. This level of automation results in an interesting relationship between the environment, automated decision aid, and the human operators. The three components in Figure 4 illustrate the relationships which exist when human operators are assisted with an automated decision aid, in order to make judgments about the state of the environment. The three links shown in the figure represent opportunities for comparison between the components. The first link to be discussed is between the state of the environment and the human operators' judgment: "Human Operator Judgment Performance." Human operators assess the state of the environment based on information provided, which may have an uncertain or probabilistic relationship with the state of the environment. This link represents the match between the operators' judgments and the actual environmental state, and therefore represents human operators' performance in making judgments.

Similarly, a link exists between the environment and the automated decision aid. The automated decision aid's task is to estimate the state of the environment based on available information, utilizing algorithms and techniques such as data fusion. A discrepancy is possible in this process when the automated decision aid is not designed to anticipate the possible states of the environment or any malfunctions in processing the collected information occur. That is, the automated decision aid may not be completely competent in estimating the state of the environment, which consequently may have an effect on the human operators' judgment performance when the environmental estimates are provided to them. One factor affecting the discrepancy between the true state of the environment and the estimates from the automated decision aid is how well the uncertain or probabilistic relationship between the environment and the information available to the decision aid is captured by and integrated into the automated decision aid. Therefore, the discrepancy that may exist within this link can be used to evaluate the competence of the automated decision aid; so labeled "Degree of Aid Competence."

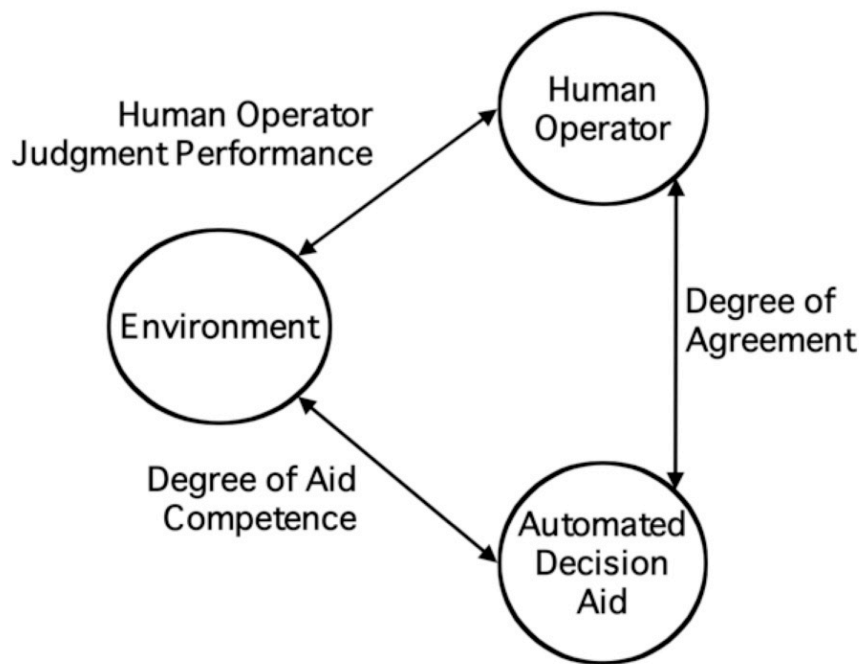


Figure 4. Three components in this study.

The last link exists between the human operators and the automated decision aid. This link represents the degree of correspondence between the judgments from the both components, denoted as “Degree of Agreement.” Due in part to the fact that the information provided to either human operators or the automated decision aid has an uncertain or probabilistic relationship to the state of the environment, and also due in part to the limit of the human’s information processing capability, there may exist discrepancy between these judgments. This discrepancy can be either very small, in which case the judgments of the human operators and the automated decision aid are mutually confirming in a certain degree, or very large, in which case the human operators must determine whether or not to integrate the automated decision aid’s estimate into the ultimate judgment. In these situations, the human operator’s trust in an automated decision aid becomes an important factor affecting the level of human operator’s reliance on the automated decision aid, which consequently may affect the overall level of human operator’s judgment performance. A contributing factor to this operator’s level of trust is the degree to which the automated decision aid provides information that is reliable, consistent, and valid. In those situations where information does not meet these requirements, the potential decrease in human operators’ trust in automated decision aid becomes a critical factor.

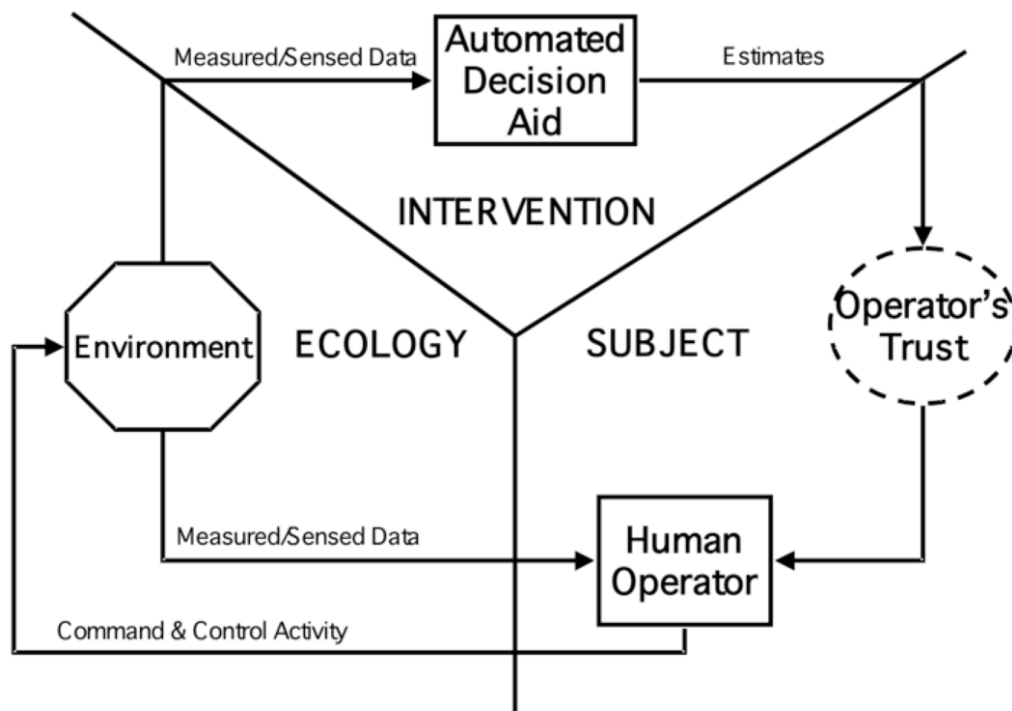


Figure 5. Flow of information among the three components in the system and the role of human operators' trust in automated decision aid in judgment and decision making.

Figure 5 shows the flow of information among the system components and activities taken by the human operator and the potential role of human operator's trust in the automated decision aid as a mitigating factor in determining the degree to which estimates from the automated decision aids are utilized by the human operators. Measured or sensed information can be provided to both the human operator and the automated decision aid so that each component can produce its own estimate. The estimates from the automated decision aid are provided to the human operator so that the operators may incorporate those estimates to a greater or lesser extent depending on the quality of the estimates, and how much the human operator puts his trust in the estimates. This is represented in the figure by placing human operator's trust in automated decision aid between the automated decision aid and the human operator. Although it has no effect on the quality of the automated decision aid's estimates, the human operator's trust in the decision aid may be a mitigating factor in the manner in which estimates from the automated decision aid are incorporated into the final operator's judgment. It is possible that when the human operators do not trust the decision aid because the decision aid has shown either unreliable or invalid estimates, they may not take the estimates from the decision aid into consideration to the final judgment. Additionally, the operators may fail to trust, and subsequently fail to use an aid because they do not understand the decision aid.

Important questions can be raised regarding the appropriateness of operator's trust in the automated decision aid. As discussed above, human trust has been defined from

various perspectives, such as sociological, and human machine interactions. Based on this literature review on the definition of trust, several system characteristics have been proposed which may influence human operator's trust in automated decision aid including aid reliability, validity, utility, and understandability. These factors should normatively influence trust in, and reliance on automated decision aid.

2.4 General Research Questions

Several classes of research questions were examined as part of this study. Although not extensive, there have been several studies on human operator's trust in automation. Based on the review of literature and conclusion drawn from the review, several research questions can be raised.

First, system characteristics of trust have been proposed in many studies from either sociological or human machine interaction setting, such as predictability, dependability, validity, and reliability. However, none of the previous studies attempted to control these variables as independent variables. Therefore, it is not certain yet which system characteristics are the most important components of operator's trust, and can affect the greatest change in trust. In prior studies (e.g., Muir and Moray, 1996; Lee and Moray, 1994), only reliability was controlled as an independent variable; other system characteristics were subjectively assessed by participants. Recall that the reliability in these studies indicates the differences between the values that the automated system actually used in controlling and the entered or reported input/output, respectively. Thus, the effect of each system characteristic of trust on the operator's trust in automated system remains unknown. Therefore, controlling the system characteristics may reveal the importance of these characteristics in contributing to different levels of operator's trust in automated systems.

Second, there has not been direct comparison made between the operator's trust in automated systems and the operator's judgment performance. While it is assumed that appropriate levels of trust lead to better performance, this has not been explicitly tested. Based on prior research in human trust in automated systems, it is hypothesized that the degree to which the human operator will rely on the decision aid can be affected by the operator's trust in the decision aid, which in turn is based on many characteristics of the decision aid, including its reliability, validity, understandability, and utility. This reliance may lead in turn to better performance.

Third, in the domain of previous research, an automatic controller was extensively used to test the effect of reliability on the operator's trust in the automation. However, none of the research has used an automated decision aid. The difference lies in the level of automation discussed at the beginning of this chapter. That is, in an automatic controller, the human operators have to engage in the automatic mode to know whether the automation performs the task well. However, an automated decision aid makes the estimates of the environmental state constantly available to the human operator. The human operator has the right to either approve or veto the decision aid's suggestion. In this way, the human operator has more chances to understand the behavior of the automated decision aid without engaging exclusively with the automated decision aid.

This may make it more difficult to find out whether the human operator actually relies on the decision aid or not. On the other hand, this may facilitate operator's calibration of trust.

However, despite the apparent difference between the automatic controller and the automatic decision aid, it is hypothesized based on the previous research that the decision aid will be trusted more when it produces estimates consistently and precisely, which are valid, and when the human operator understands the decision aid better. One study, by Bisantz and Seong (2000), examined the effect of different types of failure of automated decision aids on human performance. They identified different types of failures of the automated decision aid within the adversarial aided decision making environment, i.e., the automated decision aid attacked by the other forces. Their results showed that different types of failure induced human operators emphasizing different aspects of the decision aid.

2.5 Summary

In this chapter, literature reviews from human centered automation, trust in interpersonal relationships, and trust in human factors perspective were provided first. Based on this literature review, several interesting research issues that can be investigated to increase our understanding of human operators interacting with automated systems. To discuss these issues, general model of automated decision aids, trust, and decision making performance was provided to describe the general conception of the environment where human operators interact with automated decision aids to make judgments about the state of the environment. In this general description, the relationships between the three components were identified, which were to represent the level of performance of each judgment agent and to compare these judgment agents. These indices are important to identify the relationship between the operators' judgment performance and the use of the decision aid. Based on these two relationships, operators' trust in the decision aid can be interpreted in a normative way by linking these indices together. Discussion on this was provided in the next chapter.

Studies on trust from the interpersonal perspective and from human machine interaction perspective were provided and are summarized in the previous section. Also, the role of trust in human operators interacting with automated decision aids was described to describe the potential effect on operators' judgment and decision making performance within the general research framework. Based on this framework, several research questions were raised to understand the effects of "imperfect" automation on operators' judgment performance and trust in conjunction with automated decision aids. These questions are about the effects of different characteristics of automated decision aids on operators' judgment performance and decision making, and trust in such systems.

In addition to these research questions, another research concern in this study is to model human judgment with automated decision aids. Having the entire framework of this study in mind which is the effects of "imperfect" automated decision aids on operators' judgment performance and trust in such systems, it can be better if a model of human judgment with automated decision aid can represent a variety of characteristics of

the automated decision aid. The next chapter discusses about this modeling framework and about the development of a model of human operators' judgment with automated decision aids.

3.0 MODELING FRAMEWORK

3.1 Modeling Literature Overview

Having discussed the importance of controlling the various characteristics of systematic or human operator's trust, this chapter presents another important research goal: developing a conceptual framework from which to study human decision-making in concert with automated decision aids. It is based on several theoretical models and empirical studies described in the preceding chapter, including the Lens Model, and models studies of human trust. Specifically, this conceptual framework applies to the case in which human operators are making judgments or assessments of a situational state (e.g., assessing a level of threat in a military situation, making a medical diagnosis). To make this judgment, they have to assess informational cues as well as recommendations of state estimates from a decision aid. The human operator in this case is not mandated to rely solely on the judgment of the decision aid. It is at the human operators' discretion whether or not to rely on the situational cues or recommendations from the decision aid. It should be iterated that this is only one of several potential styles of interaction between a human and automated system, as was noted in Chapter 2.

3.2 Lens Model Framework

The Lens Model with its extensions is a modeling framework that captures both the actual environment that people make judgment about, and the manner in which humans make judgments (Cooksey, 1996). The model is the primary conceptual and methodological tool brought forward from Brunswik's (1952) idea of probabilistic functionalism. This theory indicates the functional relationship between an organism and the environment, and states that the basic of this relationship is necessarily probabilistic or uncertain in nature. As shown in Figure 6, the focal point of the modeling framework is how the information available to the human judges (the cues, labeled $X_1, \dots, X_4$ in Figure 6) is related to the actual environment state (Y_e), and how the information is used that to make judgments (Y_j) about the unknown environment. Several indices have been developed to evaluate the judges' performance level, namely environmental predictability, cognitive control, achievement, linear knowledge, and unmodeled knowledge. These parameters will be discussed in the next section.

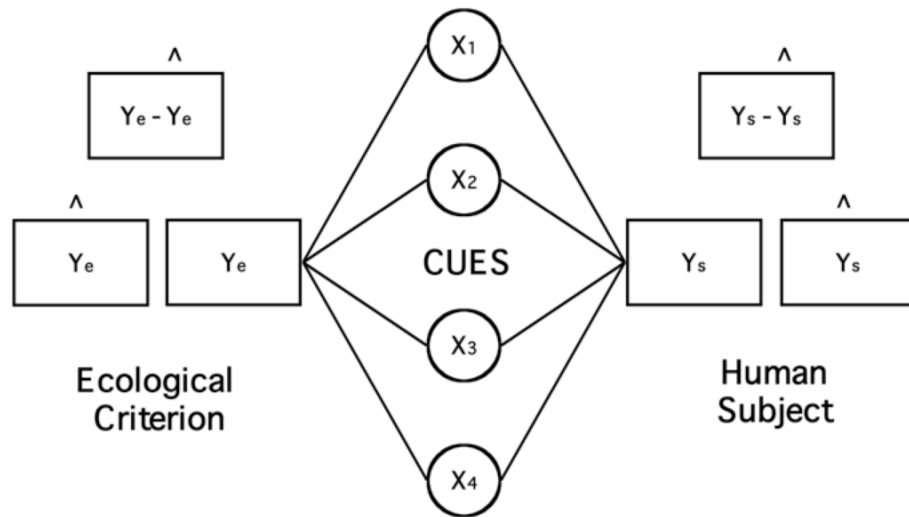


Figure 6. The generic Lens Model and double system design.

There are several different models within the Lens Model framework. These can be used to model different problem settings. Three of these model designs are discussed here to both explain fundamental concepts of the Lens Model, and to examine possible application of the model to describe human operators' judgment performance along with their use of automated systems.

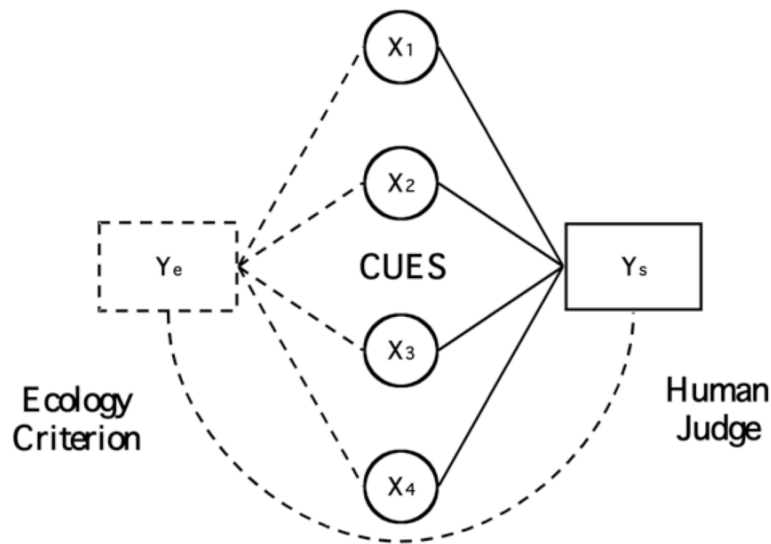


Figure 7. Single Lens Model system design

First, the single system design is the simplest design methodology for studying human judgment. As shown in Figure 7, a set of cue profiles (from X_1 to X_4) representing the unknown ecological criterion (Y_e) is available for human judge to process. The ecological criterion is shown in dotted line because for this case, its value is unknown. For this reason, it is not possible to know how well the human judge's judgments correspond to the ecological criterion. Once the judge makes judgments based on the informational cues, the types of information that can be captured to evaluate the judgment policy are the weighting scheme applied to the cues, the nature of the function form relating each cue to the judgments made, the degree of cognitive control which the judge exercises over the application of the policy, and finally the judge's organizing principle for integrating cue information. In Smith's (1975) study, he concerned planning for the future of a major U.S. university. The overall goal was to obtain views by various faculty member and academic deans. Therefore, it was not possible to know or measure what the criterion in the task ecology should or would be, thus the study was constrained to employ the single system methodology to obtain judgments of scenario desirability.

An extension of the single system design, in which the ecological criterion is known and available, is called double system design, shown in Figure 6. The double system design represents the typical Lens Model paradigm when a judge's cognitive system is explicitly compared to a known task or ecology system. Therefore, the double system design overcomes the important limitation of the single system design, in that the interrelationships between the subject's cognitive system and the task system can be examined. This type of design was applied to represent how people learn to make judgments about a criterion through acquiring an understanding of the properties of a judgment task, so called multiple cue probability learning (MCPL). Bisantz (1997) used

the double system design to represent a judgment task to determine the identity of an aircraft given a number of informational cues.

Next is the general n-system design. Figure 8 depicts an n-system Lens Model design where multiple judges make judgments about the environment based on an identical set of cues. The n-system design can be used to model a situation where the study focuses on a comparison of the judges. Basically, the n-system design is the logical extension of double system design including multiple judges. Typically, the ecological criterion is unknown, or not of interest to the researcher. Hammond & Adelman (1976) used the n-system design methodology to model a dispute between the Denver City Council and the Denver Police Department regarding the type of ammunition the police should employ in their handguns.

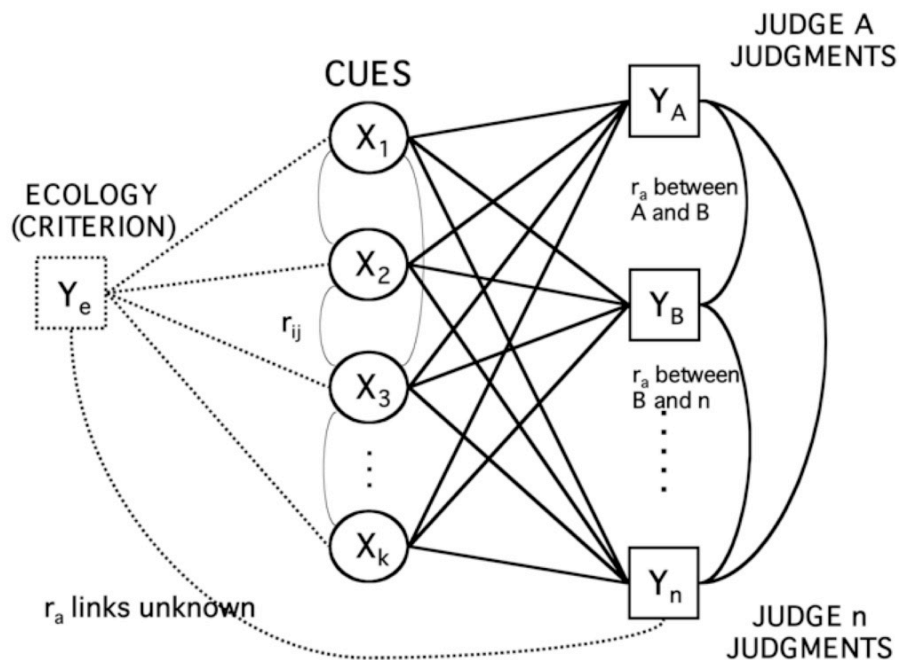


Figure 8. n-system Lens Model design.

While not frequently employed in judgment research, hierarchical judgment design is another extension of the Lens Model in which there exist multiple layers of the judgment process between the ecological criterion and the human judge. Figure 9 depicts a representation of a hierarchical system design with two layers of judgment processes and a multiple sub-domains of cues. First level judgments are made based on each subset of cues. These judgments in turn become cues for the next judgments, and so on until it reaches the final judgment. Thus, the cues in the middle between the first-order cues and the final judgment become both judgments based on the lower set of cues as well as cues for the next higher judgments. Smith (1975) used the hierarchical system design for investigation of planning judgments where alternative futures for a major university were evaluated. The overall goal was the mix of six planning area emphases desired by various faculty members and academic deans. The six areas were further subdivided into differing sets of objectives which could vary in their extent of contribution to each area.

So far, four different modeling designs in the Lens Model framework have been reviewed. A discussion on applying the Lens Model to the current research issue will be provided later in this chapter. Briefly, a Hybrid Lens Model is created by integrating the characteristics of the n-system design and hierarchical Lens Model design. The Hybrid Lens Model is based on the n-system design in which the ecological environment is known and different judgment systems make judgments about the environmental states based on a set of cues. The Hybrid Lens Model is also a hierarchical Lens Model in that the outputs from the automated system are provided to the human judges to be considered in their final judgments about the environmental states.

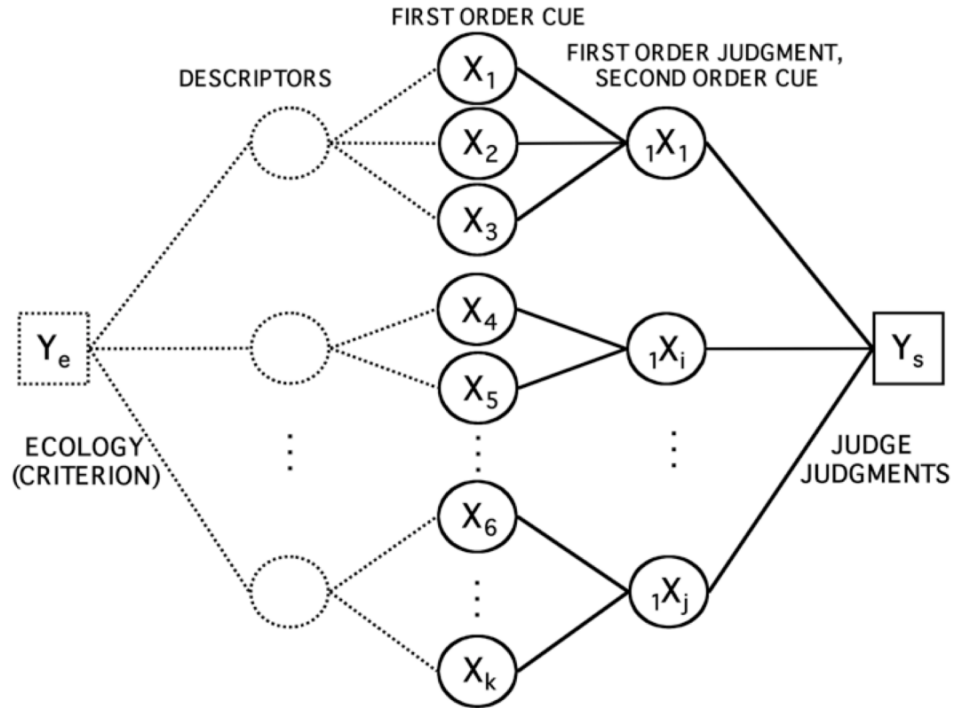


Figure 9. Hierarchical Lens Model where, k = number of cues in total, j = number of first order judgment which is the same as the number of second order cues.

3.2.1 Comparing Systems: The Lens Model Perspective

Cooksey (1996) provides an extensive discussion on the Lens Model frameworks and their associated parameters, which can be used to measure and capture people's judgment and decision-making policies. The focus in this section is on these statistical indices which arise out of this Lens Model conceptualization. The initial discussion will consider the Lens Model as implemented in the double-system design, as shown in Figure 10. This will lead to the discussion on the statistical indices of the Hybrid system design that was used in this study.

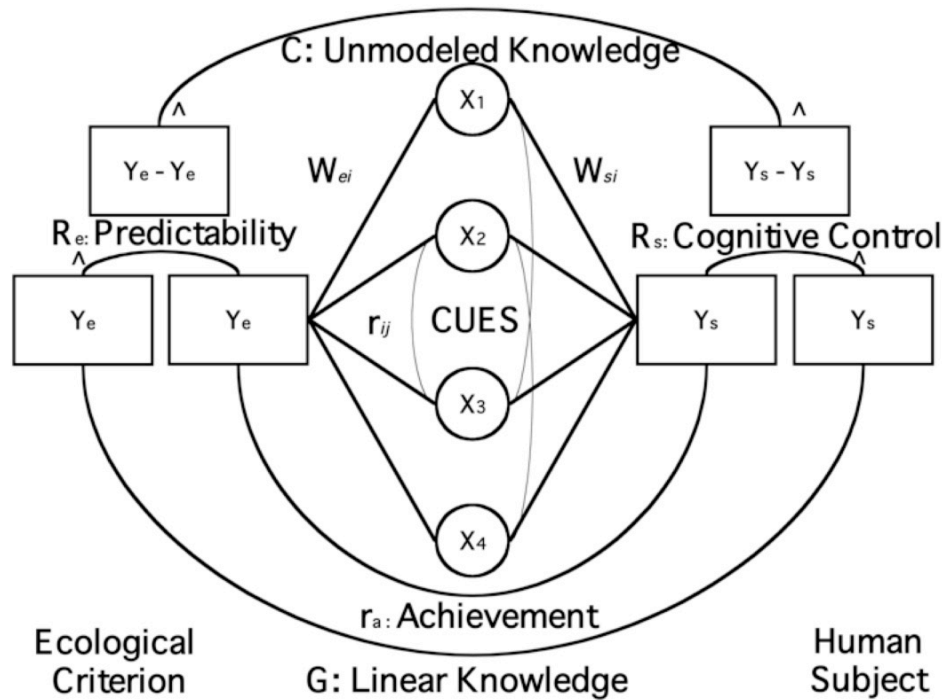


Figure 10. The Lens Model with the statistical parameters superimposed.

3.2.1.1 The Standard Lens Model Equation

Lens Model parameters can be computed based on the collection of data that includes actual judgments, environmental criteria, and cue values. The following parameters indicated on Figure 10 can be computed:

Achievement: Achievement (r_a) indicates the level of performance to the extent which the participant's judgments are in accordance with the environmental criteria. This is calculated by correlating the two fundamental components, the actual environmental criteria and the actual judgments. This correlation represents the extent to which judgments and criterion values agree. A high level of achievement represents the situation where environmental values and judgments are close together.

Cognitive Control: Once the judgments are made, multiple regression analysis between the set of cue information provided and the judgments made can be used to generate estimates for the judgments. The correlation between the actual judgments and the estimates of those judgments is used as an index to represent the degree to which the participants are able to implement their knowledge or judgment policies into the judgment task. While this index of cognitive control (R_s) is somewhat similar to the degree of consistency in executing the policy, Hammond *et al.* (1975) claimed that there existed a technical distinction between cognitive control and consistency. When the same judgment profiles whose set of cue information is identical are repeated within a set of trials, index of consistency indicates the extent to which operators can make the same judgments over those repeated profiles. This can be very critical in a situation where any devices providing information to the operators are consistently incorrect. In a military domain for example, suppose that the informational cues measured by a group of sensor

consistently indicate that the unknown objects are hostile whose true identities are otherwise. Cooksey (1996) summarized a procedure to statistically separate the level of consistency from cognitive control, which will be discussed in detail in the next Chapter.

Environmental Predictability: When the true states of the environmental criterion are known or of interest to the research, one can examine the relationship between the set of cue information provided and the environmental criterion, termed environmental predictability (R_e). The concept and the method to calculate the environmental predictability are same as calculating cognitive control of a subject. Environment predictability measures how well the set of cue information corresponds to the environmental states. Basically, this index represents how predictable the set of information is in terms of the environmental states.

Linear Knowledge: Often called G , linear knowledge measures how well the predictions of the model of the human judge correspond to predictions of the model of the environment. That is, this index shows how well the linear component of the human's judgment corresponds to the linear component of the environment. A high level of linear knowledge indicates that both models are similar and therefore predictions will be more likely to coincide. It suggests that that human judge captures the linear structure of the environment very well. The Linear Knowledge, G , is mathematically defined as the correlation between predicted criterion values and predicted judgments

$$G = \text{Corr}(\hat{Y}_e, \hat{Y}_s) \quad \text{Equation 1}$$

Unmodeled knowledge: Often called C to indicate configularity, unmodeled knowledge captures correspondence between systematic variations in both models which are embedded within the residual terms for each model. Therefore, this index captures the degree of non-linearity to the extent that both models share the same non-linear components. However, interpretation of unmodeled knowledge is somewhat different from the previous interpretation of linear knowledge. A high value for unmodeled knowledge indicates that human judge applies the knowledge of predictive but unmodeled relationship in the environment with reasonable accuracy. Note that all of these Lens Model parameters the parameters from the regression and correlation analyses.

Unmodeled Knowledge, C , is defined as the correlation between the residuals of the environment and of the human judgment model.

$$C = \text{Corr}((Y_e - \hat{Y}_e), (Y_s - \hat{Y}_s)) \quad \text{Equation 2}$$

Hursch, Hammond, and Hursch (1964) derived the first Lens Model Equation (LME) which provided the integrative statistical formulation needed to accomplish the person-ecology comparison task. Later, Tucker (1964) modified this equation which simpler formulation has become the standard form of the LME used today. The LME is

$$r_a = GR_eR_s + C\sqrt{(1 - R_e^2)}\sqrt{(1 - R_s^2)} \quad \text{Equation 3}$$

where

r_a : achievement, R_e : environmental predictability, R_s : cognitive control, G : linear knowledge, & C : unmodeled knowledge.

In Equation 3, a judge's achievement is a function of two multiplicative components. The first component of the equation is termed the linear component and represents the component of the achievement correlation which can be attributed to the explicit linear modeling. The second component of Equation 3-3, originally called configural component, and later termed the unmodeled component, reflects the presence of the unmodeled knowledge (C).

The equation incorporates an interesting point which the achievement can be calculated solely based on the linear knowledge, task predictability, and cognitive control, if the unmodeled component can be assumed to be null.

In addition to the basic Lens Model parameters, an index relevant to this study is the degree to which a judge makes judgments in a consistent manner. Consistency defined as the degree to which a judge makes consistent judgments on the profiles that have the same set of informational cues. It sounds similar to the index of cognitive control. Hammond *et al.* (1975) claimed that a technical distinction between cognitive control and consistency with respect to the interpretation of R_s should be made. This distinction between two parameters will be discussed in the next chapter in more detail because it provides important control over the independent variables in this study.

Cooksey (1996) explained the procedure for statistically separating cognitive control from consistency by using the lack of fits test by Draper and Smith (1966). First, it is necessary to test whether consistency must be separated from cognitive control. Draper and Smith (1966) provided a procedure, called lack of fits, which could identify whether the policy model accounts for all the systematic variation in the individual's judgments. The test for lack of fit is basically whether the total variance of the individual's judgments (S_Y^2) is greater than the total variance of individual's judgments for the repeated profiles (S_e^2). If lack of fits is not demonstrated (meaning that the total variance of the individual judgments for the repeated profiles is equal to or greater than the total variance of the individual's judgments), then the assumption of Y_j and $\hat{Y}_j$ coinciding is tenable and one can use R_s to index both cognitive control and consistency. Otherwise, then R_s only measures cognitive control and one needs a separate index for consistency.

3.3 Lens Model of Human Operators with Automated Decision Aids

Recall that in the previous section, the type of automated decision aid of concern in the present research was indicated to be one in which the human operator, as well as the automated decision aid, both have access to informational cues and produce judgments. The human operator is free to incorporate the output of the automated decision aid into their judgments. Based on this description of the automated decision aid and its relationship with the environment and the human operators, the Lens Model can be applied (shown in Figure 4). Figure 11 shows a Lens Model which includes the judgments produced by the automated decision aid (ADA), the human operator's judgments in conjunction with the automated decision aid, denoted HO_{ADA} , and the human operator without automated decision aid, denoted HO . The task of the operator is

to provide estimates about the state of the environment based on the information provided, denoted as X_i . In this figure, there is a dotted arrow link (between **ADA** and **HO_{ADA}**) representing the input from the automated decision aid to the human operator. This link represents not only the informational output or state estimate provided by the decision aid, but characteristics of the decision aid itself. That is, the human operator not only considers the direct estimates from the automated decision aid, but also considers the characteristics of the automated decision aid such as its reliability, and validity. Therefore, the automated decision aid, as a whole, is considered by the human operator in determining whether or not to take the estimates into consideration in the final judgment. Also, the link is shown as a dotted line to indicate that it is the operator's decision to ask to see the estimates from the automated decision aid. The locus of authority over the decision aid's estimates is the human operator. That is, when the operator's trust in the decision aid decreases, the operator may not seek the estimates from the decision aid.

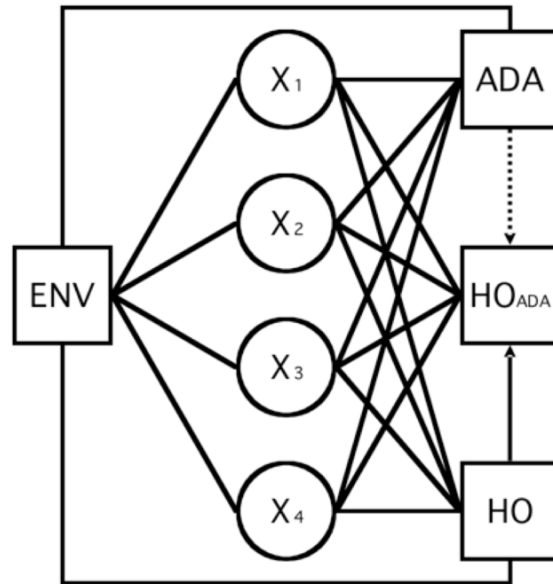


Figure 11. Application of the Lens Model. ENV: Environment, HO: Human Operator, ADA: Automated decision aid.

Figure 12 represents the Lens Model using three connected models, in order to better illustrate the comparisons between the models, annotated with several important relationships in comparing the judgment performance. While the set of information cues, which are provided to the automated decision aid and the human operator, are the same, as shown in the top and middle Lens Model in Figure 12, the estimates from the decision aid become an information cue in the Lens Model of the human operator acting in conjunction with the automated decision aid. This adds another cue in the bottom Lens Model in Figure 12. It is important to emphasize that while the model assumes that the same set of information is available to both the human operator and the automated decision aids, the model itself makes no claims regarding the degree to which or manner in which those informational cues are utilized by either of the two judgments systems. In

the most extreme case, it is entirely possible within this modeling framework for the automated decision aid to rely on a completely different subset of the available cues than the human operator. More likely, each relies on some overlapping set of cues, to different degrees. The evaluation of the manner in which cues are utilized by either is in fact an empirical question, measured through the Lens Model parameters.

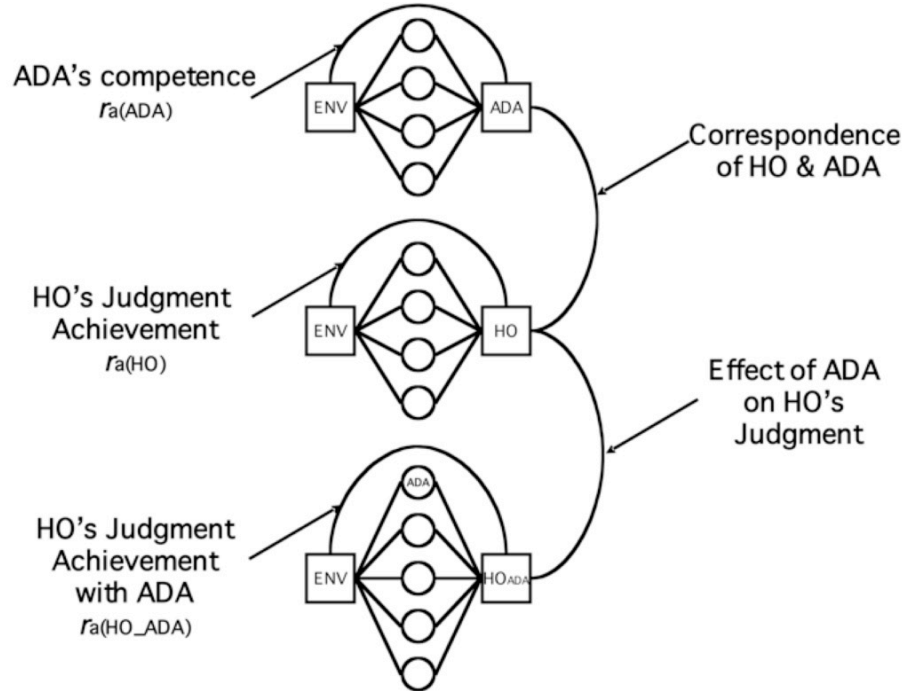


Figure 12. Connected three Lens Models of automated decision aid, human operators' judgment with or without the automated decision aid showing the correspondence and the effect of ADA on the judgments.

The arcs in each model connecting both sides of the system components refer to general achievement indicating the level of performance of each judgment agent. First, the arc in the top of Figure 12 represents the degree to which the automated decision aid can estimate the environmental state. That is, the arc represents the degree to which the automated decision aid's estimates are correlated with the environmental state. These arcs are computed by correlating true environmental state and the judgment of the right side of each Lens Model, i.e., the estimates from the automated decision aid in this example.

Second, comparing the decision aid's estimates with the human operator's judgments can provide important information to the extent which both judgment agents are similar to each other in estimating the environment. It is labeled here as the correspondence between the human operators and the automated decision aid.

Third, when the estimates from the automated decision aid are available and provided to the human operators, the level of human operators' performance may be affected. This is indicated by the arc between the human operator and the human operator with the automated decision aid, denoted **HO** and **HO_{ADA}**, respectively, in Figure 12.

These Lens Models may be able to illuminate important characteristics of human operators' trust in the decision aid and their resulting strategies in considering the decision aid's estimates. For example, assume that the automated decision aid produces accurate estimates of the environmental states. This should result in high level of performance of the human operator with the decision aid, provided that the human operator trusts the decision aid's estimate and decides to rely on it. In addition, if the human operator is able to make appropriate judgments, the level of correspondence between the human operator and the decision aid will be high and also all the levels of performance will be high.

In this section, an overall conceptual framework to study human operators' trust in automated decision aids and judgment and decision-making performance within the context of the Lens Model has been discussed. Important relationships regarding human operator's trust in the automated decision aid, and the indices from the Lens Model established have been briefly discussed. The following section examines components of each of the models in detail. Further details on each Lens model will be discussed in the following section.

3.4 Details on the Lens Models

A general discussion of the conceptual model of human operators' judgment and trust in automated decision aid has been provided in the previous section. The following section describes the three related Lens Model and their associated parameters in detail.

3.4.1 Lens Model of Human Operator's Judgment

The model of a human operator's judgment of the state of the environment is a prototypical example of the application of the Lens Model framework. In this model, shown in Figure 13, human judgment agent makes a judgment (Y_{HO}) based on the informational cues provided from the environmental state (Y_{env}). Predictions of the environmental state and the operator's judgment ($\hat{Y}_{env}$ and $\hat{Y}_{HO}$, respectively) can also be found based on their relationships with the informational cues. The model captures the human operator's judgment performance ($r_{a(HO)}$), the ability to implement the mental model of judgment (cognitive control, denoted as R_s), and the components of judgment performance which can be captured in a linear model (G_{HO}) and components of judgment which can not be captured in a linear model (C_{HO}). Additionally, R_e , environmental predictability, represents the linear predictability of the environment. Also, $W_{i(HO)}$, the operator's weighting scheme on the informational cues, can be used to characterize the operator's apparent strategy in making the judgment.

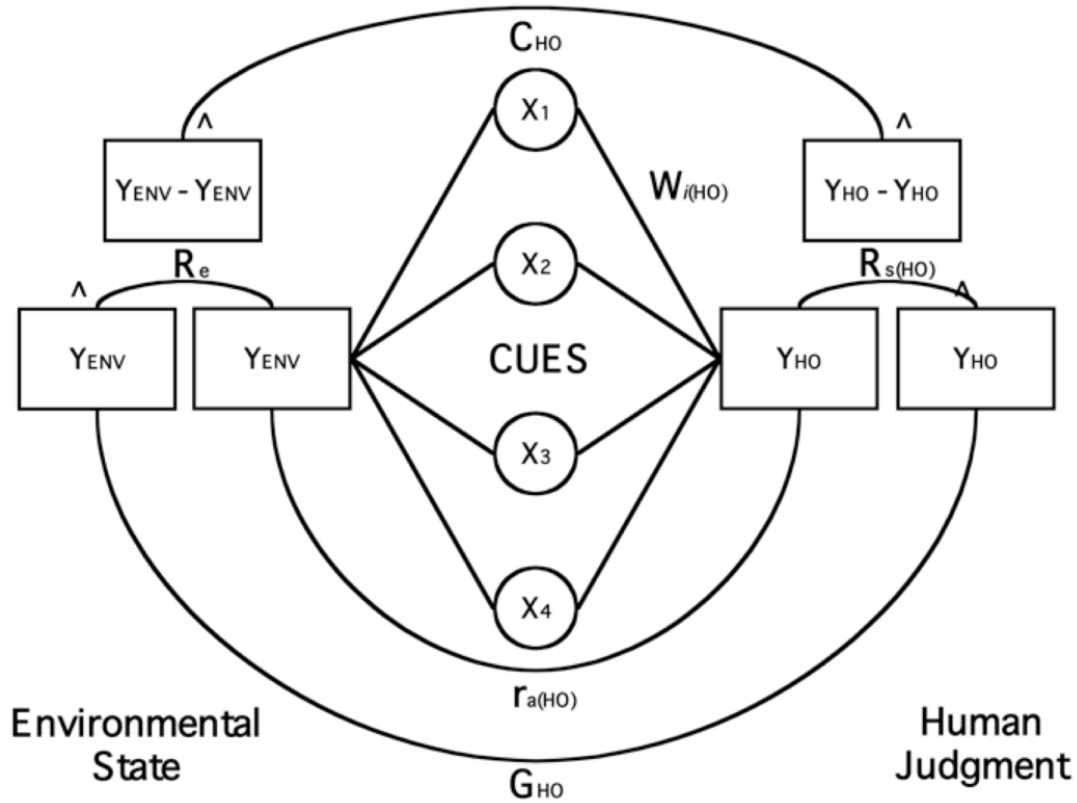


Figure 13. Lens Model of human operators' judgment without automated decision aid.

Table 6 explains the Lens Model parameters within the context of Figure 13. Since the Lens Model of human operators' judgment of the environmental state is a typical example, the description is very straightforward in accordance with the definitions of those Lens Model parameters. For example, the achievement parameter, denoted $r_{a(HO)}$, indicates the degree to which the environmental states are in accordance with the human operators' judgment strategy calculated by correlation model. Since descriptions of the Lens Model parameters are provided in the previous chapter, they will not be discussed in detail in this section.

Table 6. Description of the Lens Model parameters of human operators' judgment without automated decision aid. LMP: Lens Model Parameters

LMP	Denoted	Description
W_i	$W_{i(HO)}$	HO's weighting scheme
r_a	$r_{a(HO)}$	HO's achievement; judgment performance
G	G_{HO}	HO's knowledge about the environment
C	C_{HO}	Unmodeled knowledge
R_e	R_e	Environmental predictability
R_s	$R_{s(HO)}$	Cognitive control

3.4.2 Lens Model of Automated Decision Aid Performance

Applying the Lens Model framework to the automated decision aid is similarly straightforward (See Figure 14). Here, the state of the environment is presented itself via the cues (X_i) shown in the middle, which the automated decision aid uses to produce the estimates. This set of informational cues is the same as those in the Lens Model of human operator's judgments. The automated decision aid calculates the estimates of the environmental state that will be available for the human operators to consider and to render the judgment.

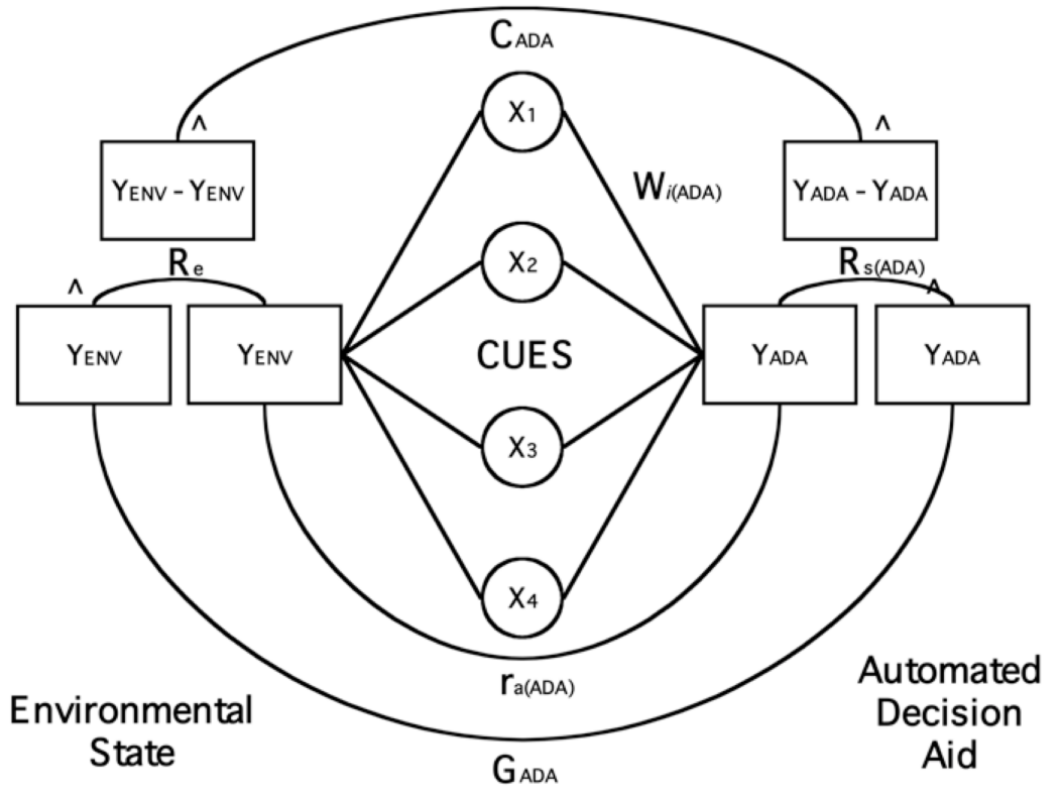


Figure 14. Lens Model of automated decision aid.

When the Lens Model is applied to model the automated decision aid's performance estimating the environmental state, the Lens Model parameters take somewhat different meanings. In this case, the Lens Model parameters can be used to evaluate various aspects of the automated decision aid. Table 7 shows the Lens Model parameters with its description of each parameter when it is applied to the automated decision aid. For example, the parameter to represent the human operators' performance, achievement denoted as r_a , becomes an index to evaluate the decision aid's performance. That is, the index represents the degree of ability or competence to which the automated decision aid can estimate the state of the environment. Also, it is interesting to note that another parameter, linear knowledge (G_{ADA}), can be used to evaluate the automated decision aid performance and competence. Generally, linear knowledge in the Lens Model describes the degree to which the predicted model of the environment and the predicted human operator's knowledge implementation are in accordance with each other. However, when it is used to describe an automated decision aid, this can represent

the knowledge that is implemented in the automated decision aid. Because this parameter represents the degree of correlation between the predicted environments and the predicted automated decision aid's knowledge base, this is essentially the same as the knowledge base implemented in the automated decision aid. Thus, this parameter can be an index which describes the degree to which the automated decision aid can perform the task for situations or the states of the environment that have not been expected in the design stage. This aspect can represent the degree of the decision aid's sensitivity to a real and unforeseen environment. Similarly, unmodeled knowledge, denoted as C_{ADA} , represents the degree of knowledge not implemented in the automated decision aid. Therefore, both parameters, linear and unmodeled knowledge, can be utilized to evaluate the degree to which the automated decision aid's knowledge base fits the actual environment.

The parameter typically known as cognitive control, denoted as $R_{s(ADA)}$, is also interesting for two related reasons. First, this parameter indicates the degree to which the automated decision aid performs consistently in the actual environment. As indicated, cognitive control is the correlation between the actual judgment and the predicted judgment. From an automated decision aid perspective, this can be seen as how consistent the automated decision aid is in making judgments based on the information provided for the estimates. The other reason why this is interesting is that because by measuring the consistency of the automated decision aid in producing the estimates, it measures an important factor that has been hypothesized to affect the level of human operators' trust in automated decision aids: the reliability of performance of the automated system. Automated decision aids that behave consistently should be trusted more than inconsistent ones.

Table 7. Lens Model Parameters for the automated decision aid. * see text for detail.

LMP	Denoted	Description
W_i	$W_{i(ADA)}$	ADA's designed scheme
r_a	$r_{a(ADA)}$	ADA's achievement; ADA's competence
G	G_{ADA}	Knowledge that is implemented in the ADA
C	C_{ADA}	Knowledge that is not considered in the ADA
$\hat{Y}_e$	$\hat{Y}_{env}$	Predicted environment
R_e	R_e	Environmental predictability
R_s	$R_{s(ADA)}$	Cognitive control (Consistency*)

3.4.3 Lens Model of Human Operator's Performance with Automated Decision Aid

When an automated decision aid is introduced to support human operators' judgment and decision-making, one way to consider the automated decision aid is as another informational cue for the human operators to consider rendering the final judgment. Because the automated decision aid integrates the information measured from

the state of the environment, it may be weighted more than other information cues. However, the extent to which the human operator relies on the automated decision aid, or any other available information cue, is an open question. For instance, as indicated above, trust in the automated decision aid may impact an operator's decision to rely on it.

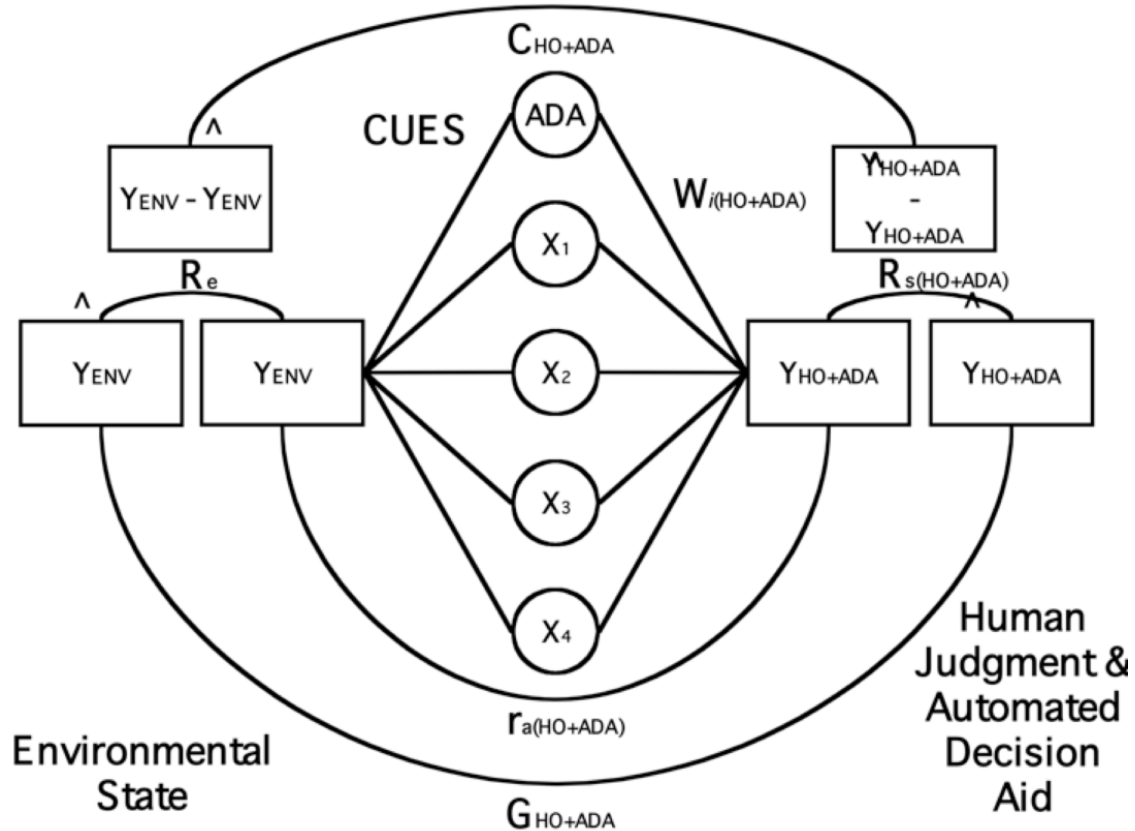


Figure 15. Lens Model of human operator's judgment with automated decision aid.

Figure 15 shows the Lens Model of human operators with automated decision aid. In this figure, the automated decision aid is added between the environment and the human operator, which means that it is considered as an informational cue similar to other information cues. In a judgment and decision-making framework, an automated decision aid can be thought of as another cue that is available to human operator to consider, but with a weight that reflects the human operator's reliance in the automated decision aid. As will be discussed later, reliance on the automated decision aid may increase as the operator trust increases. However, it is not necessarily true that the automated decision aid will be emphasized more than other informational cues.

Table 8 shows description of the Lens Model parameters of the Lens Model of human operators with automated decision aid.

Table 8. Lens Model parameters for human operators' judgment with automated decision aid.

LMP	Denoted	Description
W_i	$W_{i(HO_ADA)}$	Weighting scheme by HO with ADA
r_a	$r_{a(HO_ADA)}$	Composite achievement
G	G_{HO_ADA}	Composite linear knowledge
C	C_{HO_ADA}	Composite Unmodeled Knowledge
R_e	R_e	Environmental predictability
R_s	$R_{s(HO_ADA)}$	Composite cognitive control

3.5 Discussions on Models

Separate discussions on each Lens Model have been provided. Comparisons can be made to describe the relationships between the components which are shown in Figure 4. The relationships between the automated decision aid or the human operator and the environmental state have been discussed through the Lens Model framework. Recombining the three separate models makes the relationships between them explicit as well as illustrating the nature of the model as a combination of n-system and hierarchical Lens Model.

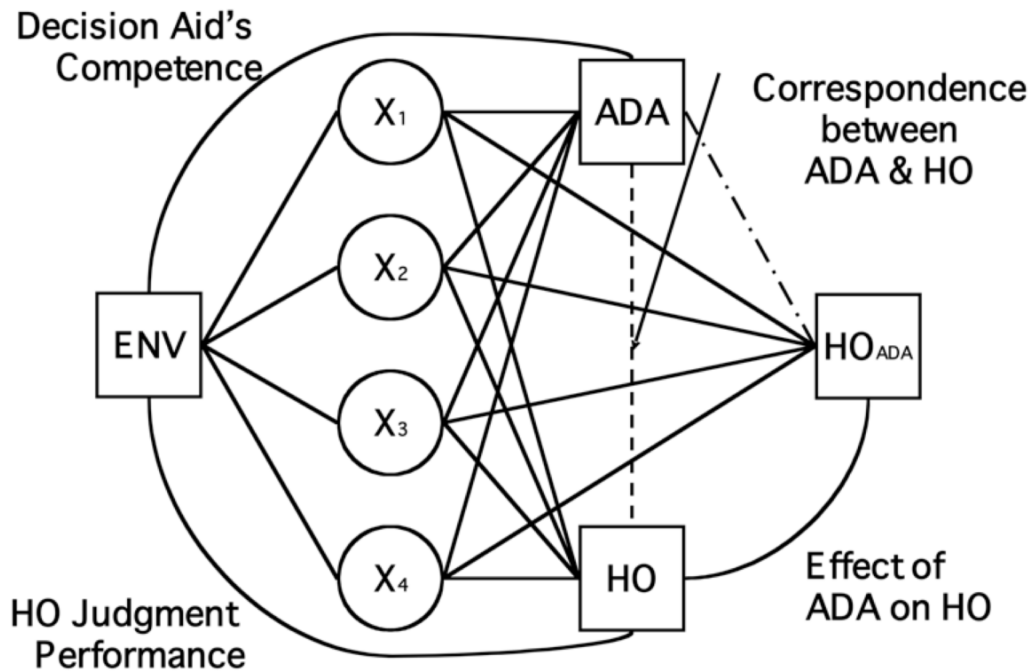


Figure 16. Hybrid Lens Model of human operators' judgment with automated decision aids.

Figure 16 depicts an overall Hybrid Lens Model of human operator's judgment with the automated decision aid. In an n-system Lens Model design, multiple judgment agents exist, and their judgment performance can be compared against each other. The model shown here is partly an n-system design in that the information cues are provided to the two judgment and decision-making agents: the decision aid and the human operator. Like one of the purposes of the n-system Lens Model design, comparing the judgments between the two judgments and decision-making agents can be accomplished.

In a hierarchical system design on the other hand, judgments of the first level agent become cues to the second level judgment agent. This model has an aspect of a hierarchical system design in that the human operator and the automated decision aid can be seen as making independent judgments, which are then combined to produce an aided judgment. Therefore, this is a hybrid model, containing aspects of both n-system and hierarchical model.

Figure 16 includes several measures representing various relationships existed within the domain of interest. First is the human operators' judgment performance, represented by the arc between the environmental criterion and the human operators, denoted "HO Judgment Performance," which can be measured by achievement in the Lens Model term. Additionally, the decision aid's competence, represented by the arc between the environmental criterion and the decision aid, can also be measured by correlating the two. Finally, the effect of automated decision aid on human operators' judgment performance, represented by the arc between the ADA and HO, can be measured by also correlating the automated decision aid's estimates and the human operators' judgments. Methodologically, this can be achieved by employing a within-subject experimental design.

3.6 Application of the Modeling Framework to the Study of Trust

3.6.1 Measuring & Manipulating Reliability, Validity, & Understandability

Within the modeling framework presented here several factors previously hypothesized to affect trust can be systematically measured, and experimentally manipulated. The reliability of an automated decision aid can be measured by the regression model indicating the consistency with which the decision aid makes judgments (e.g., $R_{s(ADA)}$). Second, the validity of the automated decision aid can be measured by $r_{a(ADA)}$, labeled in Figure 4 as the Degree of Aid Competence.

Third, the understandability of an automated system, or the degree to which human operators understand why and how the system is behaving, has been identified as a contributing factor to human operator's trust. Within this modeling framework, providing an operator with information regarding reliability and validity (that is $R_{s(ADA)}$ and $r_{a(ADA)}$, respectively) may positively contribute to understandability. Note that these types of information concern the relationship between the environment, the informational cues, and the automated decision aid's estimates. This is important because such information is identical to some forms of cognitive feedback that have been shown in previous studies to increase the level of operator's judgment performance.

Some studies have investigated the role of feedback on judgment performance. For instance, Balzer *et al.* (1989) performed an extensive literature review on different types of feedback that may have potential effects on operator's judgment and decision-making performance. These include task information, cognitive information, and functional validity information feedback. First, task information includes two relationships; one relationship between the environmental states and the set of cue information, and the other relationship between the environmental states and the estimated environmental states based on the first relationship described above. Basically, ecological validity information concerns statistical relationship between the environment and set of cue information provided to or made available to human operators. Thus, the Lens Model parameters about these relationships are included in this category; environmental predictability (R_e), cue intercorrelation (r_{ij}), and ecological validities (W_{ei}). The second type of feedback is cognitive information. It is concerned with the other end of the entire system, from the set of cue information to the human operators at the other end. This type of feedback includes human subject's cue weighting strategy (W_{si}), and cognitive control (R_s). Finally, functional validity feedback includes the types of information about the relationships between the environment and the human subjects. Among the Lens Model parameters, this type of feedback includes achievement (r_a), linear knowledge (G), and unmodeled knowledge (C).

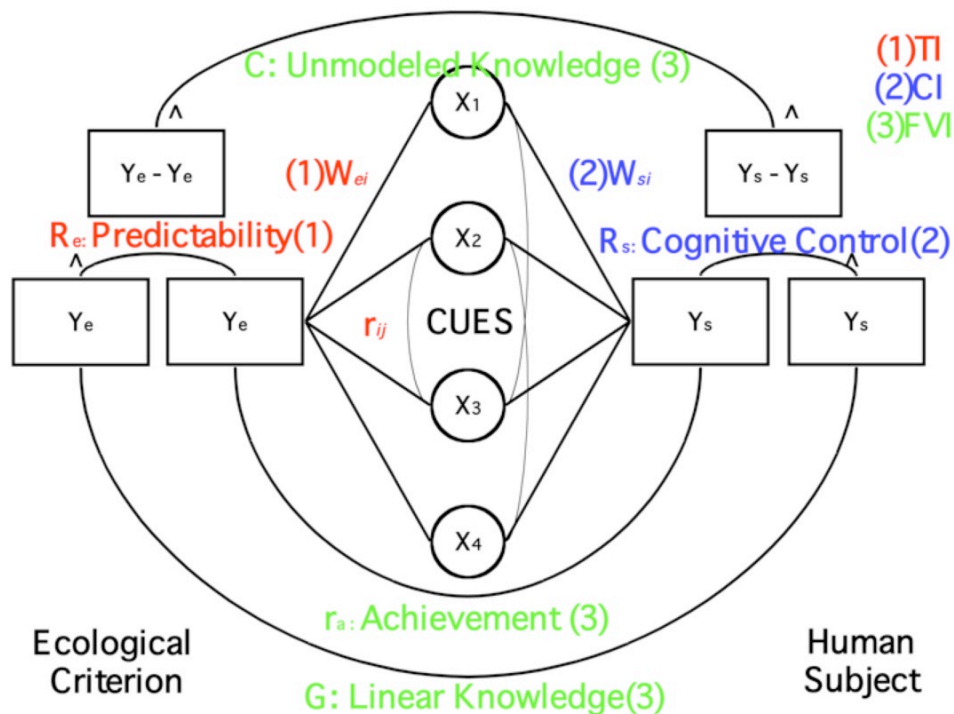


Figure 17. Lens Model with its parameters categorized by the types of cognitive feedback.

Balzer *et al.* (1992) performed an experiment to see the effects of different types of feedback on judges' performance in baseball judgment task, and concluded that the task information was the most important feature contributing the level of judges' performance. It is interesting that any further feedback provided together with the task

information did not contribute the level of performance. This has practical implications in that it can allow reducing the amount of training that judges need to perform by focusing on learning the task related information rather than concentrating on discovering their judgment policies.

On the other hand, a recent study by Gattie (2002) investigated the effects of different types of feedback on judgment performance in dental domain. Two groups of participants differing with the level of expertise in the domain were provided one of those three types of feedback. The degree of linear predictability of the environment was also controlled. He found that the cognitive information was the most important information to increase for novices in terms of the Lens Model parameters. However, experts seemed to make no difference in use of different types of feedback.

Thus, replacing the human judge in the Lens Model with one of the judgment agents in this study, the automated decision aid as was shown in the previous chapter, the Lens Model parameters transform into indices that can be used to interpret various aspects of the automated decision aid. For example, cognitive control which indicates the degree to extent which human judge is able to execute his or her knowledge into the actual judgment and decision-making task, can be interpreted as how well the automated decision aid produces the estimates consistently. Basically, those Lens Model parameters can be seen as indices that can show meta-relationships between the components. Therefore, when the same notion of cognitive feedback is applied to the automated decision aid in this study, it carries a very important link to one of the systematic characteristics of trust defined previously; understandability. That is, providing cognitive feedback of the automated decision aid can be a very efficient approach to increase the operator's level of understanding of the automated decision aid, which may lead to the development of trust.

Additionally, parameters in the model can be used to measure aspects of performance and trust. For example, the level of correspondence between the judgment of the aided human operator (with the judgment provided by the decision aid), labeled Figure 4 as the Degree of Agreement, along with the degree of similarity between the unaided and aided operator judgments, may be one possible indication of level of trust an operator has in the automated decision aid.

3.6.2 Summary of the Three Factors

Discussions on the three factors have been provided to investigate the effects of different characteristics of the automated decision aids on operators' judgment performance. The three variables were graphically shown in Figure 18 within the Lens Model framework. Validity indicated the relationship between the environmental state and the automated decision aid's estimates. This index was depicted as the connecting line between the environmental state and the automated decision aid's estimate. Reliability referred to the degree of consistency of the automated decision aid's estimates over the profiles. This index was depicted as the arrow over the automated decision aid's estimates. Finally, several types of information were identified to increase the operators' understanding of the automated decision aids, including the two factors discussed above,

and cognitive information. The cognitive information includes the weighting scheme over the informational cues, and the cognitive control of the automated decision aids that coincided with the reliability factor in this study.

Having discussed on the method of controlling these variables that were hypothesized to affect operators' judgment performance, it is necessary to examine the relationships between the three components in this study, and more importantly the relationship with human trust in automated decision aid.

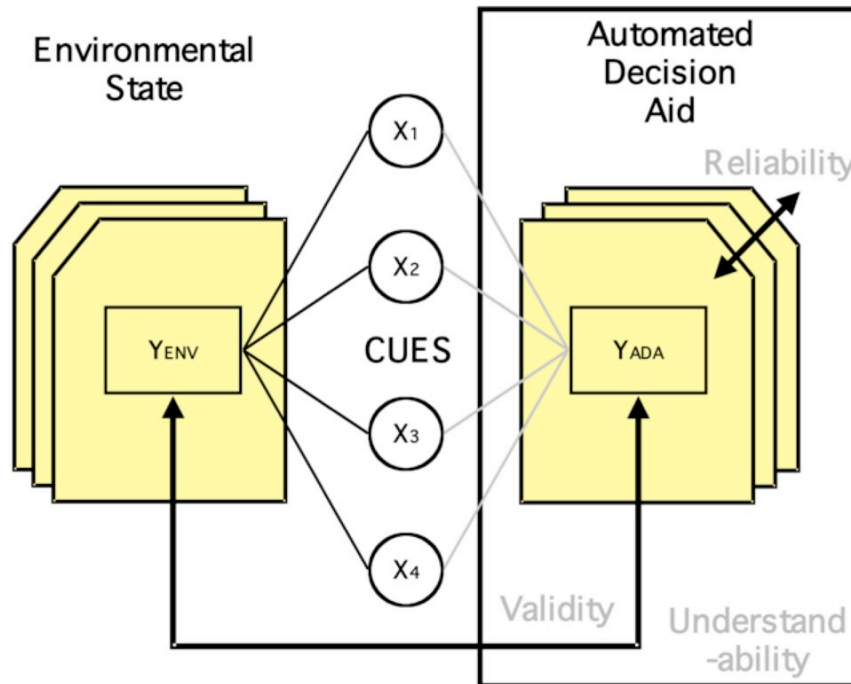


Figure 18. Representation of the three independent variables with respect to the Lens Model.

3.6.3 The Relationship between Human & Decision Aid Competence & Human Trust

Based on the degree of aid competence and the degree of agreement suggested here, several expectations and hypotheses can be made.

Figure 19 depicts how the three links shown in Figure 4 are related to each other by plotting the two degrees of similarities between the components. The degree of aid competence indicates the similarity between the environmental state and the estimates from the automated decision aid, and the degree of agreement refers to the similarity between the estimates from the decision aid and the human operator's judgment. Since the environmental state is the ultimate target value that both human operator and the decision aid attempt to identify, the degree of aid competence is an absolute value. This means that the larger the term the better. On the other hand, unlike the degree of aid competence, the degree of agreement is a relative term indicating that its value should be

interpreted depending on the degree of aid competence. For example, consider when the degree of aid competence is high (Quadrants III and IV). This means that the estimates from the automated decision aid are accurate, which also means that the human's judgment should normatively match that of the automated decision aid: a high degree of agreement (Quadrant III). In the opposite case, where the degree of competence is small, (Quadrants I and II), the degree of agreement should also be small (Quadrant I): the human, in order to make a sound judgment, should reject the advice of the aid. Therefore, while the judgment grid can represent the normative relationships between the components, by plotting the two degrees of differences on the plane also can represent the human operator's judgment behavior in accordance with the estimates from the decision aid in a descriptive way.

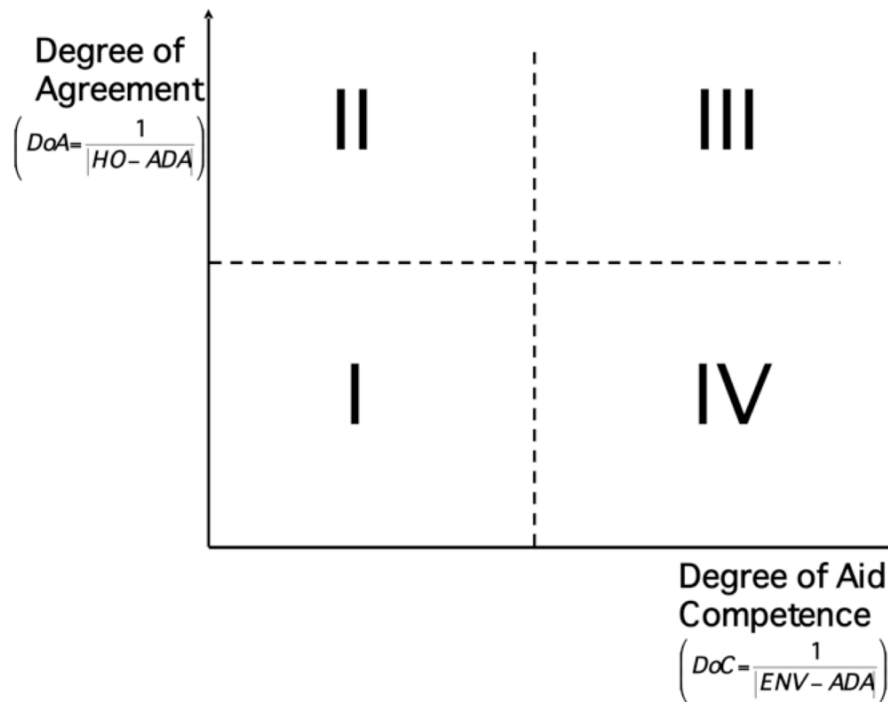


Figure 19. A template grid for analyzing operators' judgment performance relative to the automated decision aid and the environmental state, and normative level of trust.

Quadrants, II and IV, are the cases where the operators fail to calibrate their levels of trust in the decision aid appropriately. Quadrant II is indicative of the operator blindly trusting the decision aid without knowing its real competence, and/or the operator's poor ability to generate the estimate of the environmental state. This is the case where human operators "misuse" the automated decision aids. Similarly, the operator's trust in the decision aid in the Quadrant IV is low because of operator's false belief on his ability to estimate the environment. This is the case where human operators "disuse" the automated decision aids. Also, this quadrant has a special meaning when the entire sequence of the operator's judgment is plotted in this plane. Assume that the decision aid provided inaccurate estimates of the environmental state, and has performed the task well for the last few attempts. When the operator's current judgment falls into Quadrant IV, this situation can mean that the operator's trust has not been completely recovered from the

previous low level of trust. This “inertia” effect has been shown in the prior research (Lee & Moray, 1992; Lerch & Prietula, 1989) that the operator’s trust in automation does not recover to the level where it should have been. Therefore, the simple two-dimensional space can be used to estimate operators’ normative level of trust based on the two indices which represent some aspects of performance of both judgment agents in a discrete way. Also, it can be used to describe operators’ trust when they fail to perform the judgment tasks in the normative way.

Recall that causes for miscalibration can also be described in terms of this modeling framework. The *C* coefficient represents the extent to which the two systems’ policies in question employ similar and substantive, yet unmodeled, cues in the formation of their judgments. Also, the *G* coefficient represents the similarity of the linear models of the automated decision aids and the human operators exerted in the judgment and decision-making tasks. If it is assumed above that the coefficient *C* is negligible, several cases can be possible in which the agreement between two judgment agents may be low: (1) Agreement can be low because either or both agents show low level of cognitive control, which means that the two agents lack of ability to execute their judgment policy. However, low levels of cognitive control may indicate that it is possible for both judgment agents to agree in some of their judgments despite holding quite dissimilar policies. (2) Alternatively, agreement can be also low despite of the high levels of two agents’ cognitive control because the linear models of both agents are not similar.

Table 9. Operator's normative level of trust in the automated decision aid according to the quadrants in the judgment and trust grid. O: Correct, X: Incorrect judgment.

Quadrant	DoC	DoA	ADA	HO	Performance	Trust
I	Low	Low	X	O	High	Low
II	Low	High	X	X	Low	High
III	High	High	O	O	High	High
IV	High	Low	O	X	Low	Low

Cooksey (1996) discussed the first case as “false agreement,” which is shown in the top of Figure 20. With greater dissimilarity between the policies of the two judgment agents, the distance between the two means of judgments exerted by two agents is greater than when the policies are similar. The shaded region which is the overlapping region of two judgment distributions, describes the situation where the two agents’ judgments are in accordance with each other despite the largely different judgment policies. On the other hand when the judgment policies are similar which means higher *G*, regions of “false disagreement” (Cooksey, 1996) can be exhibited when an agent makes a judgment that would not be executed by the other judgment agent. This is shown in the bottom of Figure 20. The areas are those that cannot be covered by the other judgment distribution.

Hammond & Brehmer (1973) demonstrate that the results of false agreement or of false disagreement can be quite serious, and lead to distrust and an increased level of conflict between the judgment agents. That is, these two indices can be utilized as

supporting evidence to interpret judgment agents' trust in automated decision aids in conjunction with the trust grid discussed previously. Judgments that fall into Quadrant II, and IV in Figure 19 may display these two indices of inter-system conflict and the consequential level of a judge's trust in the automated decision aid, depending on the level of the coefficient G . Again, this can be very useful indices to link human judge's judgment performance and trust in automated decision aids.

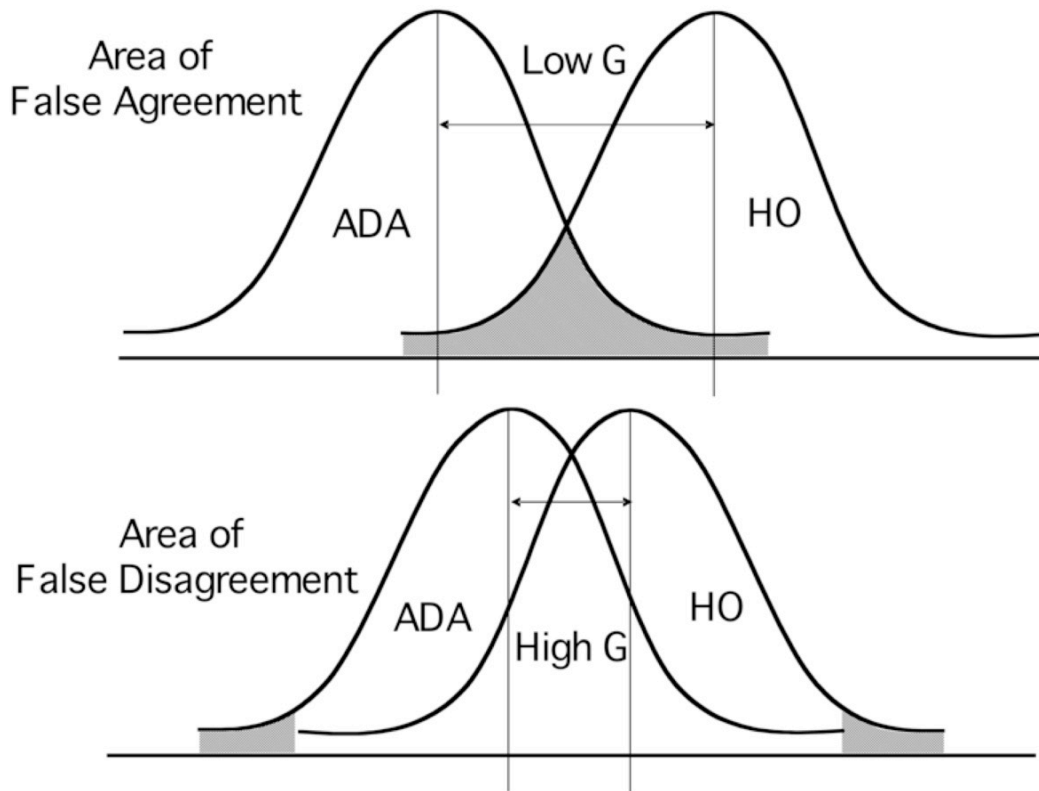


Figure 20. Regions of False agreement and False disagreement when comparing two judgment agents (Adapted from Cooksey, 1996).

4.0 METHODS

4.1 Introduction

Research findings from the previous studies in human operator's trust in automation were used to generate a set of general research questions. An experiment was designed to investigate the issues raised in Chapter 3. This experiment used an aircraft identification task, in which participants were asked to identify unknown objects on a display based on the information provided. There are several reasons why the aircraft identification task was selected. First, the task challenges human operator's cognitive ability to integrate a set of information to identify the unknown environmental state and compare it to an automated decision aid. Second, the environment has an uncertain probabilistic nature between the set of information provided to human operator and the environmental state. Third is related to the levels of automation. This task can provide the human operator with the final authority to approve or veto an automatic system based on the previous behavior of the automatic system. An automated system can perform this task independently of the human operator, providing an opportunity to examine the automated system's behavior. Finally, current technological development to integrate the information automatically to support human operator for judgment and decision-making, called data fusion technology, has been proposed for similar tasks.

In this aircraft identification task environment, estimates from a decision aid were provided, in order to characterize the relationship between the decision aid's characteristics and the operator's trust in the decision aid. A set of three independent variables was manipulated to control the various aspects of the automated decision aid and to see their effects on operator's trust in the decision aid and the level of performance: decision aid reliability, validity, and understandability.

4.2 Independent Variables

An experimental framework was established and described in the previous chapter. These independent variables have been selected to manipulate various aspects of the automated decision aids, in order to assess the degree to which they alter a person's reliance on the decision aid. The independent variables were selected based on the previous research on human operators' trust in automation. In this section, these independent variables will be discussed conceptually, and operationalized in the context of a decision aid which provides probabilistic estimates regarding a situation assessment.

4.3 Development of Experimental Testbed

A simulated bomb-detection task was selected as a testbed to investigate various issues around human automation interaction in terms of human trust and self-confidence. A Visual Basic® based stand alone program was written.

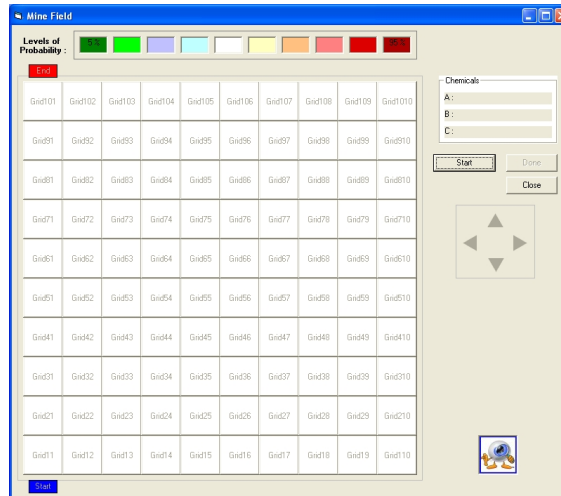


Figure 21. The initial screen of the experimental testbed built in this study.

Figure 21 shows 10X10 grid simulated landmine field. At the upper right corner, indicated as “chemicals”, displays the amount of chemicals the automated decision aid detected once the operator moves the ADA to a grid. These three chemicals were selected based on the extensive search on possible combinations of important chemicals to make landmine. For safety purpose, the names of the chemicals were not displayed on the screen. Also, the color bars located at the top of the display indicates the probability of landmine on that grid that the ADA is located currently. The operator was asked to reach to the end point as soon as possible without detonating the landmine (wrong decisions).

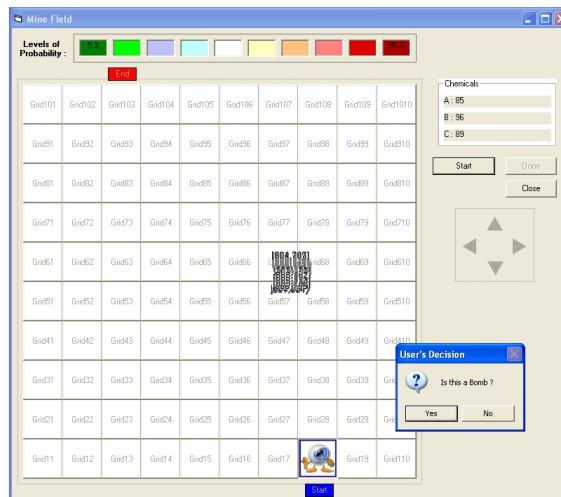


Figure 22. The second screen asks the operator's decision (first decision) based on the detected chemicals displayed on the upper right corner.

The start (blue box at the bottom) and end points (red box at the top) were randomly selected for each trial. Once the program starts, the operator moves the ADA to the first grid, which displays the levels of chemicals detected on the upper right corner. The operator is pre-trained to understand the relationship between each chemical and the possibility of the presence of a landmine. The operator is asked to make the first decision

solely based on the information about the chemical composition. Figure 22 shows this step.

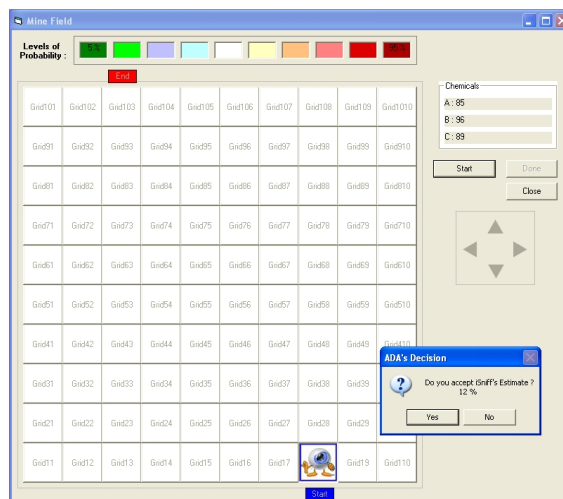


Figure 23. The third screen asks the operator's second decision after being informed from the ADA with the probability of landmine in the grid.

The operator's first decision, shown in Figure 22, brings the next screen, Figure 23. This screen displays the probability that the ADA produces based on the same chemical composition that the operator already was given. When the ADA's estimate is similar to the operator's assessment and so the first decision, the operator is asked to accept the ADA's estimate. Otherwise, the operator is asked to reject the ADA's estimate by clicking "No" button.

Finally for the decision, the operator is asked again to make the final decision on the presence of the landmine on the grid, shown in Figure 24. The previous two steps or decisions were designed to investigate any changes that the operator could have made after the operator receives the ADA's estimate.

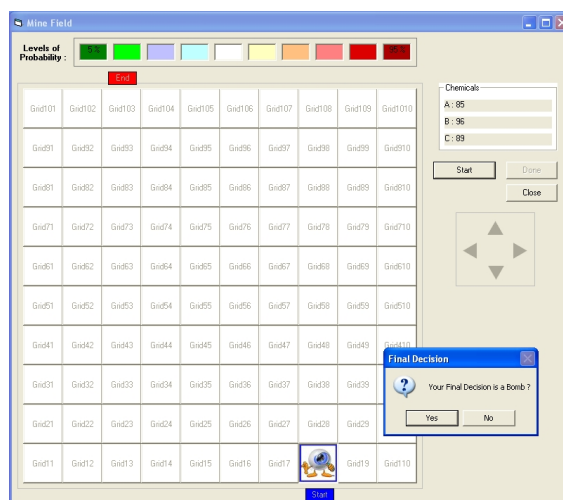


Figure 24. The fourth screen asks the operator's final decision.

Additionally, Figure 25 below shows the screen asking the operator's input on the level of self-confidence on the final decision made. This screen is designed to investigate how self-confidence plays a role on making the final decision. In the grid, three letters were displayed along with the ADA's estimate in percentage, i.e., B/N/N 12%. The first letter (B in the above example) indicates the operator's first decision (without any input from ADA). The second letter (N) indicates the operator's acceptance of the ADA's estimate (Figure 23), while the third letter (N) indicates the operator's final decision (Figure 24). Once the operator finishes the self-confidence rating, the operator is given with the feedback that indicates whether the decision was correct or not.

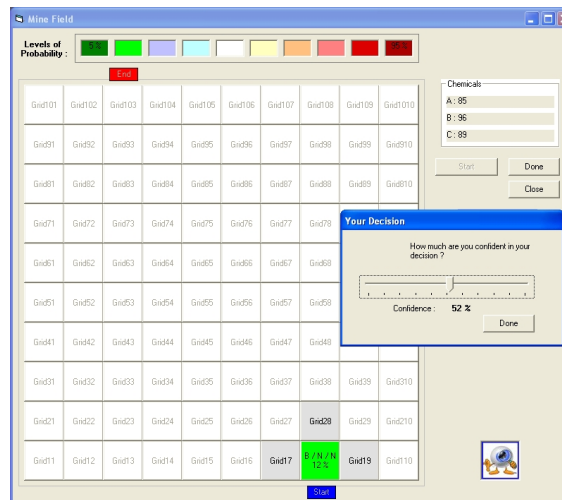


Figure 25. The fifth screen asks the operator's level of self-confidence on the final decision made.

Next, once the operator finishes with the current grid, potential or adjacent grids that the operator can move the ADA are highlighted in gray. There are four adjacent grids to any grid in this display (up, bottom, left, and right). However, for the starting point, there are only three adjacent grids highlighted, shown in Figure 26.

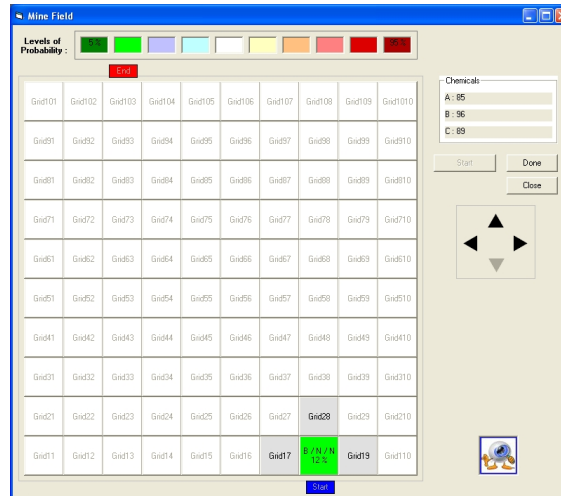


Figure 26. The next screen shows the possible grids that the operator can move once the operator made the final decision on the previous grid (grayed grids).

Figure 27 shows the path from the starting to end point. This figure is generated only for demonstration purpose. This screen allows the operator to see how the operator moved through the landmine field in order to develop a better or same (if demonstrating a good level of performance at the first trial) strategy for the upcoming trials.

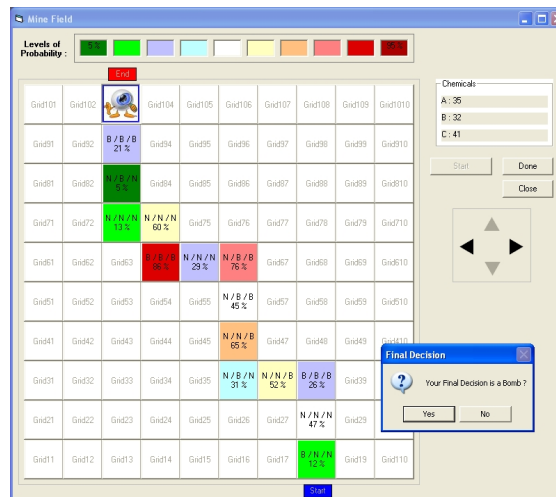


Figure 27. The screen showing the path that the operator went through. Colors indicate the level of probability that the ADA made assessments for the grid. This figure is for demonstration only.

Finally, two screens are given to rate the operator's trust (Figure 28), and importance of the trust characteristics (Figure 29). Figure 28 shows the trust questionnaire which consists of 12 questions (Llinas, Bisantz, Drury, Seong, & Jian, 1998). The operator is asked to rate one's level for each question based on the entire session. This will provide a useful way to investigate the dynamics of operator's trust on ADA.

The last screen, Figure 29, asks the operator's input on rating of importance of the trust characteristics discussed in the previous chapter. This screen is designed to provide a better understanding about the trust continuum for the potential extension of this study.

Figure 28. The trust questionnaire screen. The operator is asked to fill out the twelve questions.

Figure 29. The final screen asking to rate the importance of the characteristics to produce the final level of trust.

5.0 REFERENCES

- Balzer, W. K., Sulsky, L. M., Hammer, L. B., & Sumner, K. E. (1992). Task information, cognitive information, or functional validity information: Which components of cognitive feedback affect performance? *Organizational Behavior and Human Decision Processes*, 53, 35-54.
- Barber, B. (1983). *The Logic and Limits of trust*. New Brunswick, NJ: Rutgers University Press.
- Beth, T., Vorcharding, M., & Klein, B. (1994). Valuation of trust in open network. In D. Gollman (Ed.), *Computer Security - ESORICS 94, Third European Symposium on research in Computer Security* (pp. 3-18). Brighton, UK: Springer-Verlag.
- Bisantz, A. M. (1997). Modeling environmental uncertainty to understand and support dynamic decision making., Georgia Tech University, Atlanta, GA.
- Bisantz, A. M., & Seong, Y. (2001). Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics*, 28, 85-97.
- Brunswick, E. (1952). *The conceptual framework of psychology* (Vol. 1). Chicago: University of Chicago Press.
- Crocoll, W. M., & Coury, B.G. (1990). Status or recommendation: Selecting the type of information for decision aiding. *Proceedings of the Human Factors & Ergonomics Society 34th Annual Meeting*, 2, 1524-1528.
- Cooksey, R. W. (1996). *Judgment Analysis: Theory, Methods, and Applications*. New York: Academic Press.
- Deutsch, M. (1958). Trust and suspicion. *Journal of Conflict Resolution*, 2, 265-279.
- Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (1999). *Automation reliance on a combat identification system*. Paper presented at the Human Factors & Ergonomics Society 43rd Annual Meeting.
- Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462-492.
- Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37(2), 381-394.
- Fitts, P. M. (1951). *Human engineering for an effective air navigation and traffic control system*. Columbus, OH: Ohio State University Research Foundation.
- Gefen, D. (2000). E-commerce: the role of familiarity and trust. *Omega*, 28, 725-737.
- Glover, S. M., Prawitt, D. F., & Spilker, B. C. (1997). The influence of decision aids on user behavior: Implications for knowledge acquisition and inappropriate reliance. *Organizational Behavior and Human Decision Processes*, 72(2), 232-255.
- Grounds, C. B., and Wiley, D. (2001). *Alternative methods of mitigating automation distrust impacts*. Proceedings of the Human Factors & Ergonomics Society 45th Annual Meeting, 1728-1732. Santa Monica, CA.
- Hanes, R. M., and Gebhard, J. W. (1966). Information requirements for the control of combat forces. In J. S. Albanes (Ed.), *Reports of the committee on vision, 1947-1990*. Washington, D.C.: National Academy Press.
- Hammond, K. R., & Adelman, L. (1976). Science, values, and human judgments. *Science*, 194, 389-396.
- Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinman, D. O. (1975). Social judgment theory. In M. Kaplan & S. Schwartz (Eds.), *Human Judgment and Decision Processes* (pp. 271-312). New York, NY: Academic Press.
- Hopkins, V. D. (1994). Human performance implications of air-traffic control automation. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Recent research and trends* (pp. 314-319). Mahwah, NJ: Lawrence Erlbaum.
- Jones, S., & Marsh, S. (1997). Human-computer-human interaction: Trust in CSCW. *Ergonomics International*, 85, 172-174.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.
- Kirlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why an "Aid" can (and should) go unused. *Human Factors*, 35(2), 221-242.

- Lee, J.D., Gore, B.F., & Campbell, J. L. (1999). Display alternatives for in-vehicle warning and sign information: Message style, location, and modality. *Transportation Human Factors Journal*, 1(4), 347-377.
- Lee, J., & Moray, N. (1992). Trust, control strategies and allocations of functions in human-machine systems. *Ergonomics*, 35(10), 1243-1270.
- Lee, J., & Moray, N. (1994). Trust, self-confidence and operators' adaptation on automation. *International Journal of Human-computer studies*, 40, 153-184.
- Lerch, F. J., & Prietula, M. J. (1989). How do we trust machine advice? In G. S. a. M. J. Smith (Ed.), *Designing and using human-computer interface and knowledge based systems* (pp. 410-419). Amsterdam: Elsevier Science Publishers.
- Mockler, R. J. (1989). *Knowledge-based systems for management decisions*. New Jersey: Prentice Hall.
- Moffa, A. J., & Stokes, A. F. (1996). Trust in a medical system: can we generalize between domains? In M. Mouloua & J. M. Koonce (Eds.), *Human-Automation Interaction: Research and Practice* (pp. 218-224). Cocoa Beach, FL.
- Moon, Y., & Nass, C. (1998). Are computers scapegoats? Attributions of responsibility in human-computer interaction. *International Journal of Human Computer Studies*, 49, 79-94.
- Moray, N. (1997). Human factors in process control. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1944-1971). New York: Wiley.
- Morphew, M. E., & Wickens, C. D. (1998). *Pilot performance and workload using traffic displays to support free flight*. Paper presented at the Human Factors & Ergonomics Society Annual Meeting, Santa Monica, CA.
- Morrison, J. G., Kelly, R. T., Moore, R. A., & Hutchins, S. G. (1998). Implications for decision-making research for decision support and displays. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 375-406). Washington, D.C.: American Psychological Association.
- Mosier, K. L., & Skitka, L. J. (1996). Human decision makers and automated decision aids: made for each other. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Application* (pp. 201-220). Mahwah, NJ: Lawrence Erlbaum.
- Muir, B. M. (1987). Trust between human and machines, and the design of decision aids. *International Journal of Man-Machine studies*, 24, 527-539.
- Muir, B. M. (1994). Trust in automation. Part I Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11), 1905-1922.
- Muir, B. M., & Moray, N. (1996). Trust in automation. Part II Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39(3), 429-460.
- Norman, D. A. (1990). The "problem" of automation: Inappropriate feedback and interaction, not "over-automation". *Philosophical Transactions of the Royal Society of London*, B327, 585-593.
- Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency". *International Journal of Aviation Psychology*, 3, 1-23.
- Parasuraman, R., & Sheridan, T. B. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics*, 30(3), 296-296.
- Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). *Cognitive Systems Engineering*. New York: John Wiley & Sons.
- Reason, J. (1990). *Human error*. Cambridge, UK: Cambridge University Press.
- Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationship. *Journal of Personality and Social Psychology*, 49(1), 95-112.
- Riley, V. (1996). Operator reliance on automation: Theory and Data. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Application* (pp. 22-27). Mahwah, NJ: Lawrence Erlbaum.
- Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors*, 37(1), 5-20.
- Sheridan, T. B. (1988). *Trustworthiness of command and control systems*. Paper presented at the IFAC Man-Machine Systems.
- Sheridan, T. B., & Hennessy, R. T. (1984). *Research and Modeling of Supervisory Control Behavior*. Washington, D.C.: National Academic Press.
- Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. Cambridge, MA: MIT Man-Machine Systems Laboratory.

- Swets, J. A. (1964). Signal detection and recognition by human observers: Contemporary readings. New York: Wiley.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185, 1124-1131.
- Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-22, 589-606.
- Wickens, C. D. (1992). *Engineering Psychology and Human Performance* (2nd ed.). New York, NY: HarperCollins.
- Wiener, E. L. (1988). Cockpit automation. In E. L. Wiener & D. C. Nagel (Eds.), *Human factors in aviation*. San Diego, CA: Academic Press.
- Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, 23, 995-1011.
- Will, R. P. (1991). True and false dependence on technology: Evaluation with an expert system. *Computers in Human Behavior*, 7, 171-183.
- Woods, D. D. (1988). Coping with complexities: The psychology of human behavior in complex system. In L. P. Goodstein & H. B. Anderson & S. E. Olsen (Eds.), *Tasks, Errors and Mental Models* (pp. 128-148). NY: Taylor and Francis.
- Woods, D. D. (1993). Price of flexibility in intelligent interfaces. *Knowledge-Based Systems*, 6(4), 189-196.
- Woods, D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Application* (pp. 3-17). Mahwah, NJ: Lawrence Erlbaum.
- Yeh, M., & Wickens, C. D. (2001). Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors*, 43, 355-365.
- Zeide, J. S., & Liebowitz, J. (1987). Using expert systems: The legal perspective. *IEEE Expert*, 2(1), 19-22.
- Zuboff, S. (1988). *In the age of smart machine: The Future of work and power*. New York, NY: Basic Books.

STUDY 2

PHYSIOLOGICAL EVALUATION OF COGNITION AND JUDGMENT IN HUMAN SYSTEM INTERACTION

CHAPTER 1. INTRODUCTION

Advances in information and communication technology have led to a tremendous increase in the volume of data that can be collected. On one hand, these advances enhance human operators' understanding of their operational environments. On the other hand, the tremendous amount of data generated make human operators (HOs) susceptible to information overload, which can have negative effects on performance (Anderson, 2013; Peters, Srivastava, Taylor, Surana & Bullo, 2015). In operational environments where HOs do not have direct access to the environment, they access data about the state of the environment through a system such as human supervisory system that has a user interface. The designers of these systems are able to control how information is presented to HOs (Peters et al., 2015). These human supervisory systems aid decision making tasks by integrating relevant data collected from sensors or other agents, and then providing timely and pertinent estimates of the state of the environment to HOs.

The relevance of studying the relationship between judgment and decision making task properties (in particular display type) and performance in human factors research cannot be overemphasized (Pritchett & Bisantz, 2006; Kelton, Pennington & Tuttle, 2010). Design of user interfaces requires understanding of the judgment characteristics they are to support and the effect of their design on HO judgment (Pritchett & Bisantz, 2006). The cognitive fit between task type and the format in which information is presented leads to optimal decision making (Vessey, 1994; Speier, 2006). Furthermore, Hammond, Hamm, Grassia and Pearson (1987) found that higher correspondence between task characteristics and cognitive characteristics correlates in a significant way with HO's judgment accuracy. Thus, understanding judgment and decision making performance centered on the format of information display and the cognitive strategy adopted by the decision maker may help designers of human-machine systems design more effective screen displays, thereby improving judgment performance.

HOs use cognitive strategies to process information. Two theories that link and explain HOs' cognitive strategies to information displays and HO performance in judgment and decision making tasks are the Cognitive Fit Theory (CFT), and the Cognitive Continuum Theory (CCT). The CFT is a suitable theoretical framework for understanding the relationship between information displays and task types (Basole, Huhtamäki, Still & Russell, 2016). This theory has been validated in a variety of problem solving domains. The domains include decision making (Umanath & Vessey, 1994; Mahoney, Roush & Bandy, 2003), expertise management and visualization (Shaft & Vessey, 2006), Emergency Management (Shen, Carswell, Santhanam & Bailey, 2012), and online product presentation (Li, Wei, Tayi & Tan, 2015). However, according to Kelton et al. (2010), the relationship between presentation formats, the nature of tasks and processing of information is one of the least understood areas in the research on how presentation formats influence problem solving.

The CCT, an extension of the Lens Model (Brunswik, 1955) and the Social Judgment Theory (Hammond & Stewart, 1975), is a coalescing theory for the field of

human judgment and decision making (Cooksey, 1996). The CCT posits that multiple cognitive modes lie on a continuum between pure intuition and pure analysis, with quasi-rationality lying in the middle. The CCT is based on five premises. The fifth premise, called Structural Counterparts Premise (Hammond, 1980) attempted to relate the cognitive continuum to brain function, was later dropped. Cooksey (1996, pp 25) writes that "...the dropping of the Structural Counterparts premise may have been premature and should probably be reinstated to give impetus to research on specific brain correlates of judgment." Do intuition and analysis cognition have structural counterpart in the brain? To date, there has been little research on brain correlates to judgment.

The CCT and CFT have some similarities. According to the CFT, cognitive strategy is selected by matching information displays to tasks. When there is a match between the format of data presentation and task type, HOs are able to form the right mental representations and use cognitive strategies to fit the presented data. This fit minimizes the HO's cognitive effort leading to effective and efficient performance (Vessey, 1994). Similarly, the CCT posits that task characteristics (surface and depth) induce certain types of cognitive strategy. Correspondence between task characteristics and cognitive strategy leads to better judgment accuracy (Hammond et al., 1987).

Inferences and generalization of judgment and decision making task performance are largely made from behavioral and occasionally subjective measures. A review of CFT-based literature revealed that decision performance is mostly measured by time (response time or task completion time) and accuracy (Umanath & Vessey, 1994; Dunn & Grabski, 2001; Bacic, 2014). Likewise, measures of judgment performance in the CCT are behavioral and include achievement (correlation between a decision maker's judgments and the environmental criterion), linear knowledge (correlation between predictions made by a model of a decision maker's judgments and predictions made by the environmental model), consistency (degree to which a decision maker uses the same model from prediction to predict cue utilization), and the weighting of individual cues in making predictions. However, behavioral measures alone may not produce much information on the cognitive state of the HO (Parasuraman, 2003; Frey, Mühl, Lotte & Hachet, 2013). Physiological measures can be used alone or can be used together with behavioral measures to assess the cognitive state of the HO. In a human supervisory system where the HO's role is primarily monitoring the system, very few overt responses, for example button press, may be performed even as the HO is engaged in considerable cognitive activity (Parasuraman, 2003). In such a situation a behavioral measure such as response time (RT) provides a relatively poor sample of the mental activity of the HO. Moreover, in some cases, physiological measures may provide more information when combined with behavioral measures than behavioral alone (Kramer & Weber, 2000; Parasuraman, 2003). Physiological signals are attractive because they are continuously available and their collection does not interfere with the HO's job performance (Wilson & Russell, 2003). Physiological measures that can be used to monitor HO cognitive state include Encephalography (EEG), heart rate, eye movements, pupil size, and electrodermal measures (Wickens, Hollands, Banbury & Parasuraman, 2015). Hogervorst, Brouwer & van Erp (2014) combined and compared EEG, peripheral physiology and eye-related measures for the assessment of mental workload. They found

that EEG performed best, and that combining different sensors did not significantly improve workload assessment. EEG is well suited for human factors (Gevins & Smith, 2006).

EEG signals represent summed postsynaptic potentials of neurons firing a rate of milliseconds. Graphically, an EEG is a graph of the time varying voltage difference between an active electrode attached to the scalp and a reference electrode (Gevins & Smith, 2006). EEG may be analyzed in the time (event related potentials), frequency (spectral) or time–frequency domain. Event related potentials (ERPs) represent the brain’s neural response to specific sensory, motor or cognitive events. They are computed by averaging EEG epochs time–locked to a specific event (Fu & Parasuraman, 2006). However, ERPs fail to capture brain activity related to stimulus processing that is not time–locked to event onset (Johnson & Proctor, 2013). The frequency bands in an EEG signal are delta (less than 4Hz), theta (4– 8Hz), alpha (8 – 12Hz), beta (12– 30Hz), and gamma (30 – 80Hz) (Johnson & Proctor, 2013). During highly controlled cognitive tasks, the spectral composition of EEG changes in response to changes in task difficulty or level of alertness (Parasuraman & Rizzo, 2007). Spectral analysis involves quantifying amplitude or power of a certain frequency band. This approach is able to capture brain activity related to stimulus processing that is not time–locked to stimulus onset. Time–frequency analysis involves quantifying the spectral energy over time. It maps the 1-dimensional EEG time-domain signal into a 2-dimensional function of time and frequency.

Previous research has shown that classifiers using EEG features as inputs can be applied to the design and implementation of human-machine systems (Estepp & Christensen, 2011). Christensen, Estepp, Wilson & Russell (2012) applied three popular classifiers to EEG data obtained from participants performing a multi task with three levels of complex task difficulty across five days in one month. The three classifiers they used were artificial neural networks (ANNs) and linear discriminant analysis (LDA), support vector machines (SVMs). The classification accuracies of all three classifiers were well above chance levels. One domain where the classification of physiological features can be applied is adaptive automation. In adaptive automation, the level of automation may change during system operations (Feign, Doeneich & Hayes, 2012; Wickens, Hollands, Banbury & Parasuraman, 2015). Classification results may be used in adaptive automation to modify a HO’s task via adaptive aiding with the goal of enhancing overall performance (Parasuraman et al, 2009; Ting et al, 2010).

The first motivation for this research is the need for the design of effective decision support systems to aid today’s HOs amidst the morass of data make critical optimal decisions in a timely fashion with minimal cognitive effort. The second motivation is the need for an alternative approach to visualization evaluation other than subjective and behavioral measures. The third motivation for this research is the significant growth in the field of neuroergonomics. This has allowed researchers to study and develop new frameworks about humans and work other than an approach based solely on the measurement of subjective perceptions or overt performance of the HO (Parasuraman & Rizzo, 2007). Physiological information, along with behavioral and

subjective assessments, can be used to determine whether the modified system produces same human responses as compared to a baseline system (Kramer & Weber, 2000). Furthermore, classifiers using EEG features as inputs can be applied to the design and implementation of augmented human-machine interfaces (Estepp & Christensen, 2011). The fourth motivation is adaptive automation, similar to dynamic function allocation (Wickens et al., 2015), and the potential of using brain activity to drive an adaptive automation system (Ewing, Fairclough & Gilleade, 2016; Aricò et al., 2016; Scerbo, Freeman & Mikulka, 2003). In adaptive automation, tasks are dynamically allocated between humans and machines. The last but not least motivation for this research is adaptable user interfaces systems. These systems adapt their displays and available actions to HO's current goals and abilities by monitoring HO state, system state and current situation (Rothrock, Koubek, Fuchs, Haas, & Salvendy, 2002; Zudilova-Seinstra, 2007).

The purpose of this research is to investigate the effects of data presentation type, and task type on cognitive mode and judgment performance. Using neuroergonomics approach, I apply the CCT, along with the CFT to enhance understanding of how data representation and task type impacts decision performance.

CHAPTER 2. REVIEW OF LITERATURE

The earlier chapter served as an introduction to this research. The intention was to familiarize the reader with the motivation and purpose of the research. The aim of this chapter is to identify gaps in the literature with respect to the physiological evaluation of cognition in human computer interactions, and to build a theoretical foundation from which the research can proceed. Areas I reviewed literature include dual process theories of cognition, the cognitive continuum theory (CCT), the Lens Model with its extensions, CCT, the Cognitive Fit Theory (CFT), Electroencephalography (EEG), and Neuroergonomics.

Dual Process Theories of Cognition

Previous researchers have used dual process theories to explicate the human cognitive system. The theories are supported by much evidence in cognitive science (Evans & Stanovich, 2013), have been the focus of contemporary research (Evans & Stanovich, 2013; Kahneman, 2011; Epstein (2008); Hogarth, 2002; Sloman, 1996), and have had various labels attached to each of one of them. Theorists assume that cognitive tasks evoke two forms of processing. Evans and Stanovich (2013) refer to these as Type 1 and Type 2. Type 1 processing is autonomous and does not require working memory. On the other hand, Type 2 processing supports hypothetical thinking and requires working memory. The typical correlates of both systems are shown in Table 1. Type 1 processing is typically relatively fast, nonconscious, associative, and is independent of cognitive ability. In contrast, type 2 processing is typically slow, conscious, rule-based, and correlated with cognitive ability.

Kahneman (2003, 2011) present a two-system view of cognition, and label them System 1 and System 2. Whereas system 1 (intuition) is fast, parallel, automatic, effortless, associative, and emotional, system 2 (reasoning) is slow, serial, controlled, effortful, rule-governed, flexible, and neutral. System 1 falls to System 2 to support more detailed and specific processing when it runs into difficulty. Furthermore, System 2 is activated by an event that violates the model of the world maintained by System 1 (Kahneman, 2011).

According to Epstein (2008), humans operate two cognitive systems: an experiential system, and rational system. The operating characteristics of the two systems are presented in Table 2.

The experiential system is preconscious, automatic, nonverbal, associative, rapid, effortless, concrete, holistic, and requires minimal cognitive resources. The rational system is conscious, analytical, affect free, effortful, and requires cognitive resources.

Table 1: Attributes Frequently Associated with Dual-Process and Theories of Higher Cognition (adapted from Evans & Stanovich, 2013)

	Type 1 Process	Type 2 Process
<i>Defining Features:</i>		
	Does not require working memory	Requires working memory
	Autonomous	Cognitive decoupling; mental simulation
<i>Typical correlates:</i>		
	Fast	Slow
	High capacity	Capacity limited
	Parallel	Serial
	Nonconscious	Conscious
	Automatic	Controlled
	Holistic	Analytic
	Relatively undemanding of cognitive capacity	Demanding of cognitive capacity
	Associative	Rule-based
	Experience-based decision making	Consequential decision making
	Independent of cognitive ability	Correlated with cognitive ability

Table 2. Two cognitive systems (adapted from Epstein, 2008).

Experiential System	Rational System
1. Preconscious	1. Conscious
2. Automatic	2. Deliberate
3. Concrete: Encodes reality in images, metaphors, and narratives	3. Abstract: Encodes reality in symbols, words, and numbers
4. Holistic	4. Analytic
5. Associative: Connections by similarity and contiguity	5. Cause-and-effect relations
6. Intimately associated with effect	6. Affect-free
7. Acquires its schemas by learning from experience	7. Operates by reality principle (what is logical and supported by evidence)
8. Outcome oriented	8. More process oriented
9. Rapid processing: Oriented toward immediate action	9. Slower processing: Capable of long delayed action

Hogarth (2002) labels the two systems as tacit (or intuitive) and deliberate (or analytic), and presents a framework to understand how both systems work together. The main difference between the systems is cognitive effort. The intuitive thought is triggered effortlessly by stimuli encountered in the environment, while analytic thought is controlled. Hogarth (2002) presents three factors that affect the relative functionality of intuitive and analytic responses – the effects of “field” or “set”, the nature of the triggering stimulus, and types of cues and information. He intimates that the state of an operator at a particular time and location can impact on how he reacts to a triggering stimulus. The characteristics of the triggering stimulus will affect the extent to which

intuitive or analytic processes or both are activated. There is a trade-off that characterizes both modes, one of bias in intuitive mode versus analytical complexity in analytic mode. In a particular situation, it is essential to assess this trade-off to determine whether intuitive or analytical responses are valid. Hogarth (2002) concludes by recommending that efforts should be directed toward improving ability to make decisions in intuitive mode.

Sloman (1996) characterizes two forms of reasoning; associative system, and rule-based system. The principles of operation of the associative system are based on similarity structure and relations of temporal contiguity, while those of the rule-based system are based on symbol manipulation. Illustrative cognitive functions of the associative system include intuition, fantasy, creativity, visual recognition, and associative memory. For the rule-based system, illustrative cognitive functions include deliberation, explanation, formal analysis, verification, ascription of purpose, and strategic memory.

According to (Evans, 2012), the dual-system theories are basically the same. Some reasoning theorists (e.g., Hammond, 1996; Stevenson, 1997; Moshman, 2000) have challenged the dual-systems framework for explicating the human cognitive system, and proposed a single-system framework. They contend that the dual-system does not sufficiently accommodate the range of processes identified in studies of reasoning (Osman, 2004). Of particular interest in the present study is the cognitive continuum theory (CCT) proposed by Hammond (1996).

The Cognitive Continuum Theory (CCT)

The CCT posits that multiple cognitive modes lie on a continuum between pure intuition and pure analysis, with quasi-rationality lying in the middle (see Figure 1). Analysis cognition "...signifies a step-by-step, conscious, logical defensible process..." (Hammond, 1996, pp. 60), that is slow in data processing, high in cognitive control, has a task-specific organizing principle, and has high confidence in the data processing method but low confidence in answer (Hammond et al., 1987). Intuition cognition, on the other hand, is characterized by low cognitive control, rapid rate of data processing, low conscious awareness, weighted average organization principle, high confidence in answer but low confidence in method (Hammond et al., 1987). Quasi-rationality lies between intuition and analysis on the cognitive continuum, and is "... more or less intuitive or analytical depending upon the relative mix of intuitive and analytical characteristics demanded by the information environment" (Cooksey, 1996, p. 15).

The CCT links cognitive modes to task properties (Hammond, 1996; Cooksey, 1996), and is based on five premises summarized as:

- Premise 1: Cognition ranges from intuition to analysis on an intuition-analysis continuum.
- Premise 2: Quasi-rationality occurs between intuition and analysis.
- Premise 3: Cognitive tasks can be ordered on a continuum, based on the cognitive mode they induce.

- Premise 4: Cognitive mode oscillates along the intuitive-analytical continuum over time.
- Premise 5: Cognition is capable of recognizing general patterns and functional relations depending on task characteristics.

Hammond (1996, p.147) states that “the dichotomous view of analysis and intuition has served us badly”. Proponents of each opposing view have sought to discredit the other, accentuating strengths but failing to highlight their weaknesses. The first premise of the CCT addresses that concern. Figure 2 below shows the cognitive continuum describing characteristics of intuitive, analysis and quasi-rationality.

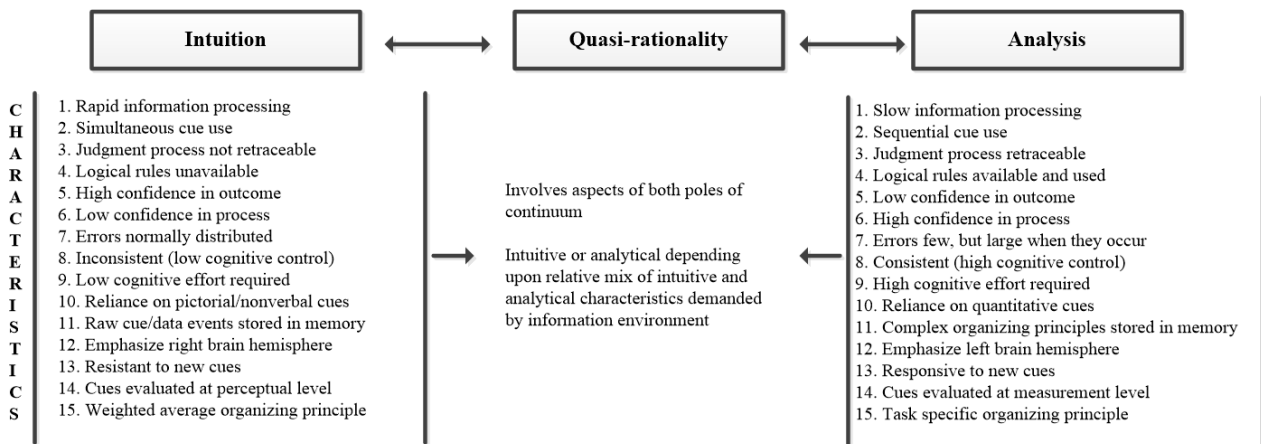


Figure 1. Cognitive Continuum describing characteristics of intuitive, analysis and quasi-rationality (adapted from Cooksey, 1996).

The second premise of the CCT postulates that there is some middle ground between the intuition pole on one end of the continuum and analysis on the either pole of the continuum. Neither intuition cognition nor analysis cognition is suitable by itself sufficiently describe all situations. This middle ground called quasi-rationality cognition is referred to as “common sense” by a layperson.

The third premise posits that tasks can be ordered on a continuum based on the cognitive mode they induce (see Figure 2). The CCT posits a Task Continuum Index (TCI), which is created by mathematically combining numerical task characteristics. Similarly, a Cognitive Continuum Index (CCI) is created by mathematically combining cognitive characteristics. The CCI operationally defines a judge’s cognitive mode. The CCI exists in parallel with the TCI. Hammond (1996, p.183) intimates that “specific representations located on the Task Continuum can be expected a priori to induce intuition, quasi-rationality, or analysis”.

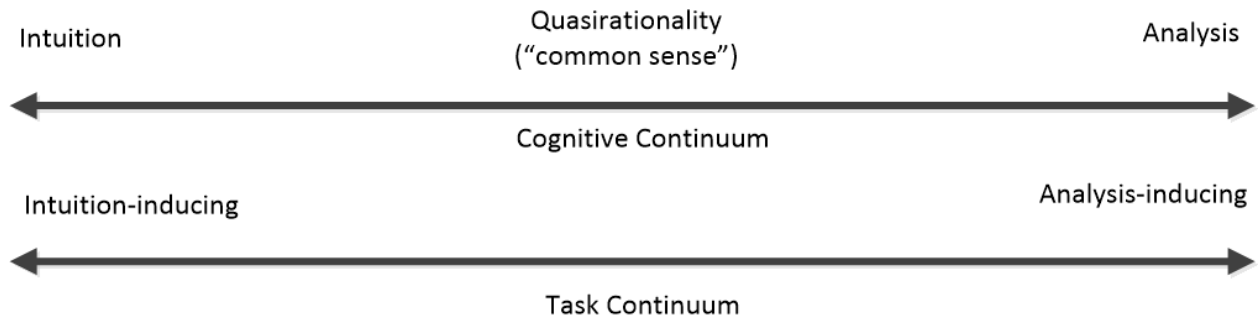


Figure 2. Cognition and tasks on a continuum.

Task Properties

The CCT differentiates between surface and depth characteristics of tasks and specifies which of them are likely to induce cognitive properties along the cognitive continuum. Whereas the depth features of the system are usually hidden from the HO, what he or she sees are the surface features which are representations of the depth features. As a result, both depth characteristics and surface characteristics need to be examined in relation to the mode of cognitive activity they induce in the HO. The task properties that differentially induce intuition and analysis are shown in Table 3 below and include number of cues, measurement of cues, distribution of cues, redundancy of cues, decomposition of task, degree of certainty of task, relation between cues and criterion, weighting of cues in environmental model, availability of organizing principle, display of cues, and time period (Hammond et al., 1987).

Table 3. Inducement of Intuition and Analysis by Task Conditions (adapted from Hammond et al., 1987).

	Task Characteristic	Intuition-Inducing State of Task Characteristic	Analysis-Inducing State of Task Characteristic
1	Number of cues	Large (>5)	Small
2	Measurement of cues	Perceptual measurement	Objective, reliable measurement
3	Distribution of cue values	Continuous highly variable distribution	Unknown distribution; cues are dichotomous; values are discrete
4	Redundancy among cues	High redundancy	Low redundancy
5	Decomposition of task	Low	High
6	Degree of certainty in task	Low certainty	High certainty
7	Relation between cues and criterion	Linear	Nonlinear
8	Weighting of cues in environmental model	Equal	Unequal
9	Availability of organizing principle	Unavailable	Available
10	Display of cues (applicable to surface conditions only)	Simultaneous display	Sequential display
11	Time period	Brief	Long

The fourth premise of the CCT introduces the element of time. Tasks that must be accomplished within a relatively short time demand intuition cognition. When there is enough time, analytical cognition may be used. Movement along the continuum may oscillate between intuition and analysis if time permits.

The fifth premise of the CCT posits that the human judge is capable of both pattern recognition (used in intuitive cognition) and of discerning functional relations (used in analytical cognition). Hammond intimates that cognition that employs functional relations is different than cognition that seeks or recognizes patterns. He states that “Pattern seeking and pattern recognition are useful cognitive activities when employed in appropriate circumstances – that is, on tasks that offer coherence that we can recognize, either through our natural capacities to do so or through our trained capacities to do so” (Hammond, 1996a, p. 198).

The Lens Model with its extensions

The Lens Model with its extensions is a modeling framework that captures both the actual environment that a HO makes judgment about, and the manner in which the HO makes judgments (Cooksey, 1996). It was originally created by Brunswik (1952) to represent the principles of probabilistic functionalism. Probabilistic functionalism captures the functional relationship between a judge and the environment, and depicts the judge-environment relationship as probabilistic relations among environmental variables (Cooksey, 1996). The Lens Model provides double, symmetric models of both the judge and the environment. The judgments made by the judge and the environmental criterion to be judged are defined as linear combinations of environmental cues. Different models or design types exist within the Lens Model framework, and these design types can be used to model different problem settings.

The single system design (Figure 3), for example, focuses on only on the judgment process. The ecological criterion is unknown, and so it is impossible to determine how well judgments corresponds to the ecological criterion.

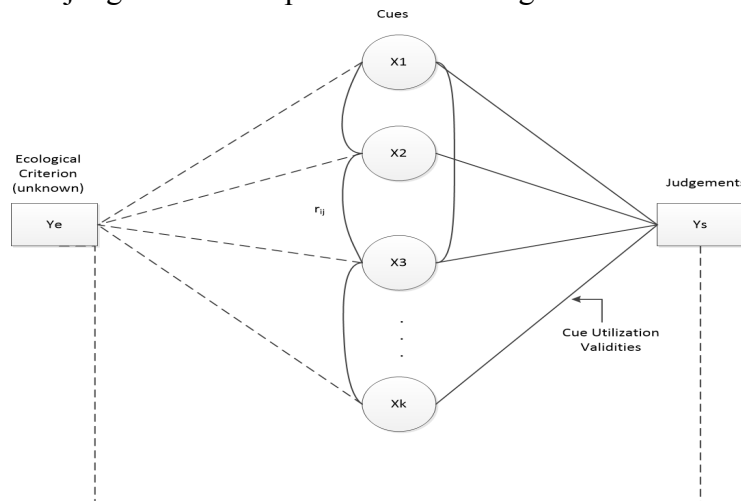


Figure 3. Lens Model representation of single system design. Adapted from Cooksey (1996).

The double system design (Figure 4) is an extension of the single system design in which the ecological criterion is known. It represents the typical Lens Model paradigm where a judge's cognitive system is explicitly compared to a known ecology system (Cooksey, 1996). This double system design has been used to describe how people make medical judgments (Wigton, 1988), educational judgments (Cooksey et al., 1996), and weather forecasting (Stewart et al., 1982).

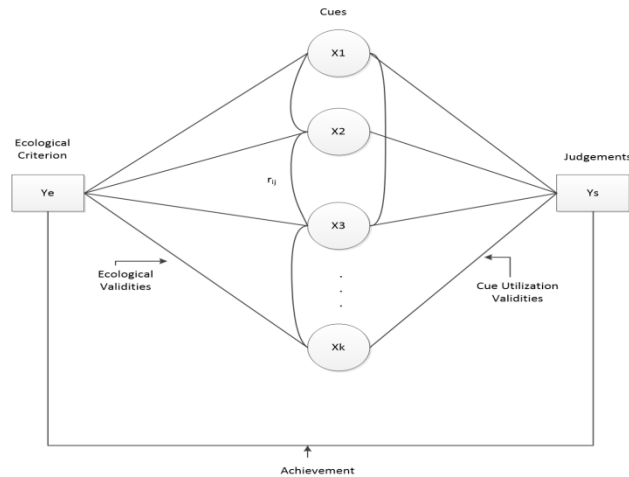


Figure 4. Lens Model representation of double system design. Adapted from Cooksey (1996).

The triple system design (Figure 5) is an extension of the double system in which two judges interact with a known task system. Comparisons are made between judge and task, and also between the judges themselves.

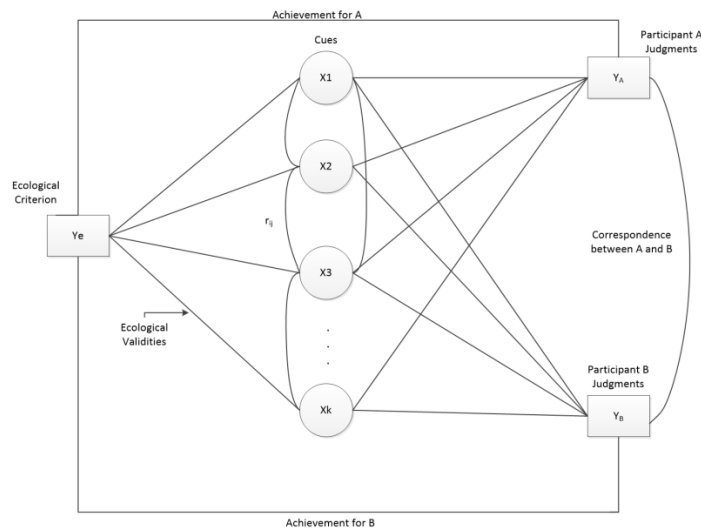


Figure 5. Lens Model representation of triple system design. Adapted from Cooksey (1996).

The n-system design (Figure 6) is an extension of the double system design to include many judges. The judges make judgments about the environment based on identical set of cues.

Usually, the task system is not known, or not of interest to the researcher.

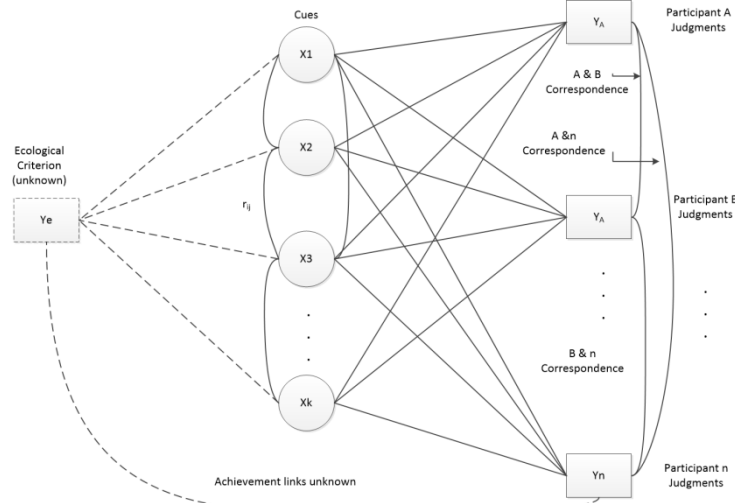


Figure 6. Lens Model representation of n-system design. Adapted from Cooksey (1996).

The Standard Lens Model Equation

The Lens Model framework and its related parameters are able to measure and capture judgment and decision making policies. The left side of Figure 7 represents the model of the environment. This describes the relationship between the ecological criterion value (e.g. good quality, bad quality) and the cue values obtainable at the time a judgment is made. The right side of Figure 7 depicts the judge's policy or strategy. It describes the relationship between the cue values and judged criterion value.

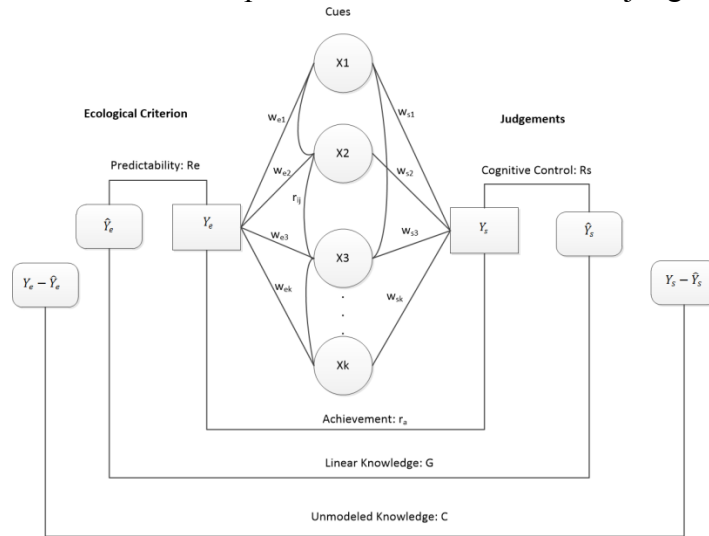


Figure 7. The Lens Model with superimposed statistical parameters. Adapted from Cooksey (1996).

In Figure 7, the judgments are related to each cue i by w_{si} , known as cue utilization validity. The pattern of cue utilizations demonstrated by a judge determines his or her judgment policy. Y_s represents judgment and is modeled as a linear combination of a set of k cues ($X_i, i = 1, \dots, k$).

$$Y_s = \sum_{i=1}^k \beta_{si} X_i + e \quad (1)$$

In Equation 1, the β_{si} s represent the weights that the judge gives to the different cues and e represents the extent to which the judge misses the actual value for Y_s he or she is trying to predict (Cooksey, 1996). Thus if $\hat{Y}_s$ denotes the cognitive model, then

$$Y_s = \hat{Y}_s + e \quad (2)$$

The correlation between Y_s and $\hat{Y}_s$ denoted by R_s , measures the cognitive control or consistency with which a judgment strategy is executed (Cooksey, 1996).

$$R_s = \text{corr}(Y_s, \hat{Y}_s) \quad (3)$$

The ecological criterion is related to each cue i by w_{ei} , known as ecological validity. The ecological validity of a cue measures how well it specifies the true state of the criterion. Y_e represents ecological criterion and is modeled as a linear combination of a set of k cues ($X_i, i = 1, \dots, k$).

$$Y_e = \sum_{i=1}^k \beta_{ei} X_i + e \quad (4)$$

In Equation 4, the β_{ei} s represent the weights that the judge gives to the different cues and e is the error term of the regression of Y_e on the X_i s. Thus if $\hat{Y}_e$ denotes the environmental model, then

$$Y_e = \hat{Y}_e + e \quad (5)$$

The correlation between Y_e and $\hat{Y}_e$ denoted by R_e , measures environmental predictability, that is, how well the value of the environmental criterion can be predicted with a linear model of the cues.

$$R_e = \text{corr}(Y_e, \hat{Y}_e) \quad (6)$$

The correspondence between judgments and the environmental criterion, called achievement, and denoted by r_a is computed using the Lens Model Equation (Hursch et al., 1964; Tucker, 1996)

$$r_a = GR_e R_s + C \sqrt{(1 - R_e^2)} \sqrt{(1 - R_s^2)} \quad (7)$$

Linear knowledge, denoted by G in Equation 8, measures how well the cognitive model corresponds to environmental model. It is computed as the correlation between predicted criterion values and predicted judgments.

$$G = \text{corr}(\hat{Y}_e, \hat{Y}_s) \quad (8)$$

Unmodeled knowledge, denoted by C in Equation 9, measures the extent to which the judgment and the environment share the same nonlinear components. It is computed

as the correlation between the residuals of the environmental model and the cognitive model.

$$C = \text{corr}(Y_e - \hat{Y}_e, Y_s - \hat{Y}_s) \quad (9)$$

The Lens model parameters given by, r_a , G , C , R_e , R_s , are described in Table 4 below.

Table 4. Description of Lens Model Parameters.

Lens Model Parameter	Description
r_a	HO's judgment performance
G	HO's knowledge about the environment
C	Unmodeled knowledge
R_e	Environmental predictability
R_s	HO's cognitive control or consistency

The linear regression model has been used to model cognitive model (the right side of the Lens Model) and environmental model (left side of the Lens Model) in most research judgment analysis research (Cooksey, 1996; Hamm & Yang, 2016). This is appropriate when the judgment or criterion is continuous. However, for studies in which the judgment or criterion is dichotomous, the general lens model equation needs to be modified. Logistic regression model may be used to model both the cognitive and environmental models (Cooksey, 1996; Stewart, 2004; Hamm & Yang, 2016).

According to Cooksey (1996, p.299), there C terms are needed for the logistic regression lens model. These terms (C_1 , C_2 , C_3) reflect the correlation between the logistic regression residual of the environment and logistic regression residual of judgments (C_1), the logistic regression prediction of the environment and the logistic regression residual of judgments (C_2), and the logistic regression residual of the environment and the logistic regression prediction of judgments (C_3). If $\tilde{Y}_e$ and $\tilde{Z}_e$ represent logistic regression prediction of the environment and logistic regression residual of the environment respectively, and $\tilde{Y}_s$ and $\tilde{Z}_s$ represent logistic regression prediction of judgments and logistic regression residual of the environment respectively, then

$$C_1 = \text{corr}(\tilde{Z}_e, \tilde{Z}_s) \quad (10)$$

$$C_2 = \text{corr}(\tilde{Y}_e, \tilde{Z}_s) \quad (11)$$

$$C_3 = \text{corr}(\tilde{Z}_e, \tilde{Y}_s) \quad (12)$$

In modifying the logistic regression lens model from the linear lens model, Stewart (2004) intimated that environmental predictability, R_e , should be replaced with $\sigma_{\tilde{Y}_e}/\sigma_{Y_e}$, ratio of the standard deviation of the logistic regression prediction of the environment to the standard deviation of the data, and the $\sqrt{(1 - R_e^2)}$ with $\sigma_{\tilde{Z}_e}/\sigma_{Y_e}$, ratio of standard deviation of the logistic regression residual of the environment to the standard deviation of the data. Similarly, the cognitive consistency, R_s , should be replaced with

$\sigma_{\tilde{Y}_s}/\sigma_{Y_s}$, ratio of the standard deviation of the logistic regression prediction of judgments to the standard deviation of the data, and $\sqrt{(1 - R_s^2)}$ with $\sigma_{\tilde{Z}_s}/\sigma_{Y_s}$, ratio of standard deviation of the logistic regression residual of judgments to the standard deviation of the data. From Stewart (2004), the logistic regression lens model is given by

$$r_a = G \frac{\sigma_{\tilde{Y}_e} \sigma_{\tilde{Y}_s}}{\sigma_{Y_e} \sigma_{Y_s}} + C_1 \frac{\sigma_{\tilde{Z}_e} \sigma_{\tilde{Z}_s}}{\sigma_{Y_e} \sigma_{Y_s}} + C_2 \frac{\sigma_{\tilde{Y}_e} \sigma_{\tilde{Z}_s}}{\sigma_{Y_e} \sigma_{Y_s}} + C_3 \frac{\sigma_{\tilde{Z}_e} \sigma_{\tilde{Y}_s}}{\sigma_{Y_e} \sigma_{Y_s}} \quad (13)$$

where $G = \text{corr}(\tilde{Y}_e, \tilde{Y}_s)$

Prior Work Using CCT

Hammond et al (1987) conducted an experiment to test the CCT. They sought to demonstrate that TCI values could be correlated with CCI values along parallel continuums. Participants were each presented with nine judgment tasks. They employed three different task surface characteristics and produced three different TCI scores ranging from low to high on a scale of 1 to 10. The three different task surface conditions were film strips, bar charts and numerical formulas. Each of the three task surface conditions was associated with three different task depth characteristics producing three different CCI scores also ranging from low to high on a scale of 1 to 10, for a total of nine task surface-depth combinations. The task depth characteristics involved judgments about highway aesthetics (an intuitive task), highway safety (a quasi-rational task) and highway capacity (an analytical task). Eight out of the eleven task characteristics presented in Table 3 were used. Hammond et al (1987) computed the TCI by determining a value for each of the eight elements. To produce a TCI value for a given task, they combined the nine sub-indices as follows:

1. Each of the sub-indices was rescaled (by linear transformation) onto a 1-10 scale, in which 1 is the most intuition-inducing and 10 the most analysis-inducing end of the scale (The rescaling reversed the scale for sub-indices 1 and 2).
2. The eight rescaled sub-indices were averaged with equal weights to produce the TCI score for the task.

Hammond et al. (1987) used four cognitive properties (cognitive control, presence of an organizing principle, error distribution resulting from the subject judgments, and the difference in confidence the participants had between their methods and their answers) to compute the CCI. They measured cognitive control by the lens model parameter, R_s . The presence of an organizing principle was measured as the difference between the coefficient of determination (r^2) of a nonlinear model of the participant's judgments and the coefficient of determination (r^2) of a linear model of the same judgments. They measured error distribution by the kurtosis of the distribution of the differences between the participant's judgments and the criterion, with a leptokurtic distribution considered to be reflective of analytical cognition. They measured the differential confidence subjectively as the difference between the participant's confidence in his or her method and his or her confidence in his or her answers. Hammond et al. (1987) expected method

confidence to be higher in analytical cognition than intuition cognition, and answer confidence to be higher in intuitive cognition than analysis cognition.

Hammond et al. (1987) tested five hypotheses:

- H1: The surface and depth task properties induce corresponding cognitive properties.
- H2: Analytical cognition will not always provide a ceiling for performance
- H3: Analytical cognition is apt to produce extreme errors.
- H4: The greater the correspondence between surface and depth task properties, the better performance will be.
- H5: Knowledge of the congruence between the surface and depth characteristics of the task is necessary and sufficient to predict performance.

They found that the task location as identified by the TCI score did induce the location of cognitive activity in the predicted order for most engineers in most conditions, but did not induce cognitive activity located at a corresponding point on the cognitive continuum for all engineers in every condition. Intuitive or quasi-rational cognition outperformed analytical cognition in 11 of 20 cases, indicating that analytical cognition does not provide a ceiling on performance. The most serious errors committed by the subject engineers in the formula task condition were in the most analysis-inducing depth task, that of judging highway capacity. Performance was found to be better when cognitive properties of a particular subject corresponded to task properties that subject faced. Simple knowledge of the congruence between surface and depth characteristics of the task was not sufficient to predict improved performance. Performance was enhanced if subjects employed cognition suited to the task properties. They concluded with three suggestions. First, experts should increase their awareness of the correspondence between task and cognition. Second, when the display features of the task are not able to be changed, then the expert should estimate the location of the task on the task continuum and adapt his cognitive activity to that indicated by the TCI score. Third, when the cognitive activity is unalterable, the task displays should be used that correspond to the cognitive activity.

Hamm (1998) followed up on the work by Hammond et al. (1996). Hamm was one of the investigators in the Hammond et al. study. Hammond et al.'s study had a molar focus and considered whether a whole task would induce intuition or analysis and whether the participant's cognition was entirely intuitive or analytical. Hamm's study had a molecular focus, and he sought to demonstrate that cognition oscillated between intuition and analysis in moment-by-moment variations while HO was engaged in a complex judgment task. He created an index of cognition, the moment-by-moment cognitive continuum index (MBMCCI) that ranged on a scale from 0 (most intuitive) to 1 (most analytical). Hamm found a pattern of mode shifts (analytic – intuitive – analytic) between intuitive and analytical cognition. He attempted to correlate his results with those of Hammond et al. for the same engineers on the same tasks. The correlation between his mean MBMCCI and Hammond et al.'s CCI for the same engineers on the same task was low.

Dunwoody, Haarbauer, Mahan, Marino & Tang (2000) tested specific aspects of the CCT. The experiment involved a multiple-cue threat assessment task. They manipulated task conditions to present two task surface conditions, and three task depth conditions.

The two task surface conditions were a pictorial display of threat cues (intended to induce intuitive cognition) and a numeric tabular display (intended to induce analytical cognition). The three task depth conditions were manipulated to include one at the intuition-inducing pole of the task continuum, one in the quasi-rational center of the task continuum, and one at the analysis-inducing pole. Their TCI was based on four factors: number of cues, the cue redundancy, the ecology cue weight and the task predictability. The cue redundancy was measured by the average of the intercue correlation coefficients. The ecology cue weight was measured by the average standard deviation of the set of β weights from the regression model of the environment. The task predictability was measured by the lens model parameter R_e^2 . Each value of these four factors was scaled to a positive value between 1 and 10. Then they were weighted and averaged to produce the TCI score (see Equation 14). This was different from that of Hammond et al. (1986).

$$TCI = \frac{((10 - cues) + (10 * (1 - \bar{r})) + (10 * \sigma_\beta) + (10 * R_e^2))}{40} \quad (14)$$

Dunwoody et al. (2000) constructed the CCI using five factors: consistency, error distribution, response rate, self-insight into policy, and differential confidence. Consistency was measured by the lens model parameter R_s . Error distribution was measured by the kurtosis of the distribution of the residuals ($Y_s - \hat{Y}_s$). Response rate was measured directly in seconds. Self-insight into policy was subjectively measured by comparing differences between measured cue utilization weights and reported cue utilization weights. Differential confidence was subjectively measured by the difference in the participants' confidence in the method they used in their judgment and their confidence in the accuracy of their answer. This was different from that of Hammond et al. (1986). Dunwoody et al. (2000) attempted to show that the score of the TCI they devised was correlated with the index of cognitive mode. However, one result was unexpected. The task that was predicted to induce a high cognitive index value (demonstrating analytical cognition) rather produced a low cognitive index value, almost identical to the task inducing the most intuitive cognition.

Cooksey (2000) compared Hammond et al. (1987)'s study to Dunwoody et al. (2000)'s in an attempt to provide guidance for future researchers. With participant selection, he noted that Hammond et al. (1987) used expert highway engineers in their study whereas Dunwoody et al. (2000) used undergraduate students. With respect to research design, Hammond et al. (1987) employed a within-subject design whereas Dunwoody et al. (2000) employed a between-subject design. According to Cooksey (2000) the between-subjects design was a weak test of the CCT. Cooksey posits that a within-subjects design provides a stronger test of CCT because the cognitive movement between modes can be traced for each judge.

Brunswik's probabilistic functionalism seeks to explain that the functional relationship between an organism and its environment. It is based on probabilistic relations among environmental variables (Cooksey, 1986). Two principles of the probabilistic functionalism are vicarious functioning and vicarious mediation (Cooksey, 1996). According to Hammond (2006) vicarious mediation refers to the environment's

way of conveying uncertain information through a variety of observable indicators. Similar levels of prediction can be achieved using different cues if those cues are intersubstitutable or correlated in nature (Cooksey, 1996). On the other hand, vicarious functional refers to the use of a variety of intersubstitutable, observable indicators to achieve a reasonable accurate inference about the matter of interest. Similar levels of functional response can be achieved using different cues provided the cues are intersubstitutable or correlated in nature (Cooksey, 1996). Stewart (2001) discusses the lens model and the work of Brunswik into detail. He explains that the degree of vicarious mediation present in an environment is a function of both cue intersubstitutability and ecological validity weights of cues. He suggests that vicarious functioning is characteristic of intuitive cognition and that the potential for vicarious functioning on the part of the subject can be predicted by examining the variability present in the vector created by the matrix product between the matrix of cue ecological validity weights and the matrix of cue inter-correlations.

The lens model equation parameter, G , is referred to as linear knowledge and it measures how well the cognitive model corresponds to environmental model. Stewart (2001, p.360) used the formula of G in Equation to investigate vicarious functioning.

$$G = \frac{\beta_e' \mathbf{R} \beta_s}{\sqrt{\beta_e' \mathbf{R} \beta_e} \sqrt{\beta_s' \mathbf{R} \beta_s}} \quad (15)$$

where β_e is a vector of standardized regression weights for the environment (ecological validity), β_s is a vector of standardized regression weights for the judge (cue utilization), and $\mathbf{R}$ is the matrix of cue intercorrelations (Stewart, 2001, p.360). The product of β_e and $\mathbf{R}$ may be used to determine how a judge uses vicarious functioning. When the matrix product $\beta_e' \mathbf{R}$ produces a vector of numbers that have low variability, then the potential for vicarious functioning is high. This means that the variability of the vector $\beta_e' \mathbf{R}$ may serve as an indicator of the degree of vicarious mediation present in the environmental conditions of a given task. This variability may also serve as an indicator of the potential for vicarious functioning on the part of the judge (and hence for the degree of intuitive cognition). Stewart (2001, p. 350) writes that "This exploration of G leads to an important insight: Opportunities for vicarious functioning depend not only on the intercorrelations among the cues, but on a combination of those correlations and the ecological validities of the cues. As far as we know, Brunswik was not aware of the importance of both cue intercorrelations and ecological validities in vicarious functioning."

Holcomb (2011) created new metrics for measuring task properties and cognitive mode, and demonstrated better support for the hypothesis that judgment performance could be improved if cognitive mode matched task properties. Previous attempts had failed to do that. He created an alternate set of TCI and CCI indices based on the work of Stewart (2001). He conducted an experiment involving a judgment task similar to the experiment done by Dunwoody et al. (2000). The potential for vicarious functioning within a task, as measured by Stewart (2001)'s variability of the matrix product $\beta_e' \mathbf{R}$ was selected to serve as an indicator of the degree of intuitive cognition possible for that task. This was a basis for the new TCI score. Holcomb (2011) called this new index Vicarious

Mediation Index (VMI) and claimed it reflects the degree of vicarious mediation inherent in the task characteristics. Similarly, Holcomb (2011) created a new CCI index called Vicarious Functioning Index (VFI). He defined VFI as the mean deviation of the vector produced by the matrix product of the cue utilization weights (β s) and the cue intercorrelations R . He investigated if there was a predictable relationship between vicarious mediation inherent in the task conditions and vicarious functioning on the part of the judge, and if matching between the VMI and VFI indices predicted task achievement better than the earlier experiments. Unlike Dunwoody et al. (2000)'s experiment, Holcomb (2011) employed a within-subject design and experienced middle school teachers to make judgments on student placement using a variety of familiar cues. He used both the VMI index and Dunwoody et al. (2000) TCI index and compared the results. Results showed a significant relationship between task properties and cognitive mode, and improved task performance when there was close correspondence between the task properties and the cognitive mode. His results showed that judgment performance can be improved when task properties and cognitive mode are considered together. However, the cues that Holcomb (2011) used for his experiment were hypothetically generated and it is unclear how they relate to the judgment task he used in his study. Furthermore, Holcomb (2011) does not explain why he used negatively correlated cues in his study, and how that influenced the results.

The Cognitive Fit Theory

The Cognitive Fit Theory (CFT) was developed to explicate how the fit between information presentation format and decision making task can influence HO's problem solving performance (Vessey & Galletta, 1991). The CCT posits that problem solving performance depends on the information presentation format and the nature of the task. Different presentation formats accentuate different problem solving strategies. In the same vein, different judgment and decision making tasks accentuate different problem solving strategies. The CCT suggests that when there is a match between information presentation format and task type, HOs are able to form the right mental representations and use cognitive strategies to fit the presented data. This fit minimizes the HO's cognitive effort leading to timely and decision making performance (Vessey, 1994).

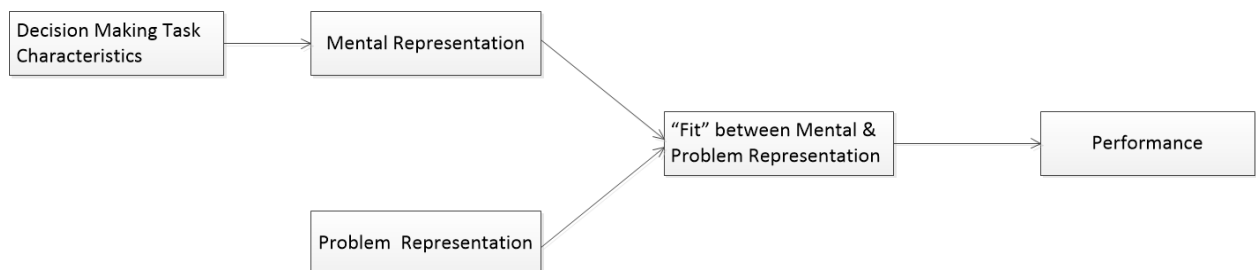


Figure 8. The Cognitive Continuum Theory (adapted from Basole et al., 2016).

In the CFT, problem representation (information presentation format) and decision making task are viewed independently (Figure 8). Information presentation may be in graph (spatial) or table (symbolic) format. Graphs and tables represent the same data in different ways. Graphs are spatial problem representations, and they facilitate

viewing the presented information without attending to the elements analytically. Graphs accentuate relationships in the data (Vessey, 1991).

Tables, on the other hand, are symbolic problem representations, and they facilitate analytical processing of the presented information. Tables do not represent data relationships openly. However, they facilitate extracting data values (Vessey, 1991).

Decision Making Task Characteristics

Einhorn and Hogarth (1981) posited that judgment and decision making tasks consist of three interrelated sub-tasks: information acquisition, evaluation, and feedback/learning. Information acquisition processes are information search and storage. Information acquisition tasks are executed directly via information acquisition processes. One type of information acquisition task is spatial tasks. In the CFT, these tasks require the HO to assess the problem area holistically and are best accomplished using perceptual processes.

The second type of information acquisition task is symbolic tasks. They require the HO to extract discrete data values and are best accomplished using analytical processes. Information evaluation tasks are basically spatial or symbolic in nature. As evaluation tasks gets more complex, the more complex symbolic tasks involve analytical evaluation, and more complex spatial tasks involve perceptual evaluation. Vessey (1991) argued that only two sub-tasks (information acquisition and evaluation) are relevant to the analysis of graph/table decisions and that the CFT does not need to consider feedback.

Speier (2006) investigated the effects of information presentation formats on complex task decision-making performance. She classified tasks by type and complexity (see Table 5). With respect to task type, she classified tasks into spatial and symbolic tasks. Spatial tasks display information as a whole and require HOs to perceive relationships in the data. Spatial tasks usually ask for the comparison of values, trends, or assessing relationships in the data. They are best accomplished using perceptual processes (Vessey, 1991, 1994). Symbolic tasks, on the other hand, require extracting specific and precise data values for solutions. Symbolic tasks typically request a specific answer in response to the task. They can be best accomplished through analytical processes.

With respect to task complexity, Speier (2006) classified tasks into simple and complex tasks (Speier, 2006). In the CFT, simple tasks require one action. Information acquisition tasks are simple tasks (Umanath & Vessey, 1994; Vessey & Galletta, 1991). Complex tasks require higher-level decision making. Judgment and inference tasks (Speier, 2006; Vessey, 1991, 1994) are complex tasks.

Simple-spatial tasks require spatial information acquisition and evaluation subtasks and have lower levels of the complexity characteristics. Simple-symbolic tasks require symbolic information acquisition and evaluation subtasks that involve a low number of information cues, processes and inter-relatedness within the task. Complex-

spatial tasks require spatial information acquisition and evaluation subtasks and have higher levels of complexity characteristics. Complex-symbolic tasks require symbolic information acquisition and evaluation subtasks that involve a large number of information cues, processes and inter-relatedness within the task (Wood, 1986).

Table 5. Classification of CFT Tasks (adapted from Speier, 2006).

		Task Complexity	
		Simple	Complex
Task Type	Spatial	Simple-Spatial Task	Complex-Spatial Task
	Symbolic	Simple-Symbolic Task	Complex-Symbolic Task

Prior Work Using CFT

The CFT has been validated in a variety of problem solving domains. Table 6 provides a summary of prior significant work including domain, data representations, and contributions. In particular, for this study, I focus on the prior work conducted on multi-attribute judgment tasks.

Table 6. Summary of prior research using CFT.

Reference	Domain	Task	Data Representation	Dependent Variables	Contributions
Vessey and Galletta (1991)	Bank account management	Spatial vs. Symbolic	Graphs vs. Tables	Time and Accuracy	Made suggestions for the extension of CFT to more complex problem solving environments
Umanath and Vessey (1994)	Bankruptcy decision-making	Low information load vs. high information load prediction (holistic task)	Chernoff faces vs. graphs vs. tables	Time and Accuracy	First study to extend CFT to judgment and decision making tasks
Tuttle and Kershaw (1998)	Employee performance evaluations	Analytic (Symbolic) vs. Holistic (Spatial)	Graphs vs. Tables	Time and Accuracy	Extended CFT to judgment strategy by providing evidence that matching information presentation to strategy improved judgment performance
Speier (2006)	Operations management	Spatial-simple, spatial-complex, symbolic-simple, and symbolic-complex	Graphs vs. Tables	Time and Accuracy	Evaluated CFT in complex task environment and provided a suggestion for more nuanced approach to task complexity
Shen, Carswell,	Emergency Management	Horizontal vs. Vertical vs. Combine	Plan view vs. Elevation view vs.	Time, Accuracy and Workload	First study to evaluate roles of decision guidance and adaptable

Umanath and Vessey (1994) conducted the first study to extend the CFT to judgment and decision making tasks. They examined cognitive load and the ability of the CFT to explain the performance of certain display formats on multi-attribute judgment task of bankruptcy prediction. They hypothesized that predictions made with graphs would be more accurate than those made with Chernoff faces. This is because whereas graphs possess integrating capabilities as well as permitting access to the underlying data, schematic (Chernoff) faces provide integrating capabilities but not access to the underlying data. Umanath and Vessey (1994) also predicted that there would be no difference in accuracy or performance between Chernoff faces and tables. This is because each of them, Chernoff faces and tables, provides only one of the two necessary types of data. Chernoff faces do not preserve underlying data, while tables do not provide integrating capabilities. Also, Umanath and Vessey (1994) expected Chernoff faces to result in faster problem solving compared with tables and graphs. They did not expect any difference in time for predictions made with tables and graphs. All the hypotheses of Umanath and Vessey (1994) were supported.

Tuttle and Kershaew (1998) conducted a multi-attribute judgment study to assess the performance of a hypothetical plant manager. Participants employed holistic and analytical strategies to solve judgment problems using either graphs or tables. Presentation format (graph and table) was a between-subject variable while strategy (holistic and analytical) was a within-subject variable.

Tuttle and Kershaew (1998) found that for the holistic strategy, graphs produced higher judgment accuracy manifested in consistent judgments than tables. They concluded that display format impacted accuracy through judgment consistency and not through judgment time.

Electroencephalography (EEG)

The human brain is enclosed in the cranium, bones of the skull which protect the brain surfaces. The scalp which is a thin cover of skin covers most part of the cranium. The cerebral cortex is the largest part of the brain and is located beneath the bones of the cranium. It is divided into four main lobes: frontal, parietal, occipital, and temporal. The frontal lobe is associated with reasoning, problem solving, planning, parts of speech, movement, and emotions. The parietal lobe is associated with perception of stimuli, recognition, movement, and orientation. The occipital lobe is associated with visual processing. The temporal lobe is associated with perception and recognition of auditory stimuli, memory, and speech. The cerebral cortex is made up of nerve cells called neurons. Many of these neurons are functionally connected to each other, and connected to other parts of the brain.

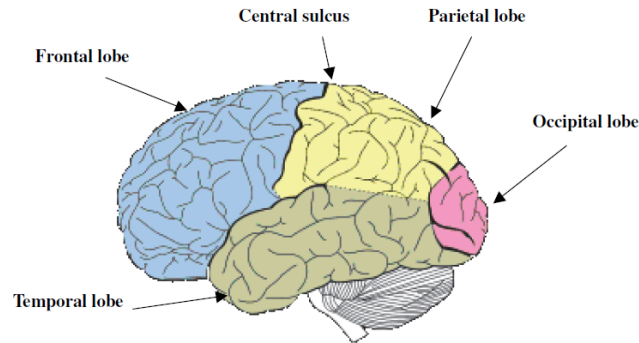


Figure 9. The Cerebral Cortex (adapted from Cheng & Hsu, 2011).

Neurons have cell body, soma, and specialized cell parts, dendrites and axons. Dendrites convey electrical signals to the soma and axons transmit information away from the soma. Neurons contain synapses. A synapse consists of a presynaptic ending that contains neurotransmitters, mitochondria and other cell organelles, a postsynaptic ending that contains receptor sites for neurotransmitters, and a synaptic space between the presynaptic and postsynaptic endings. Information exchange between two neurons is accomplished by movement of chemicals across a synapse. Neurons create action potentials, which are discrete electrical signals that travel down axons. This causes the release of neurotransmitters at the synapse. This neurotransmitter then fits into a receptor in the dendrite of the neuron that is on the other side of the synapse, the post-synaptic neuron. The neurotransmitter, when combined with the receptor, typically causes an electrical current within dendrite of the post-synaptic neuron. Thousands of post-synaptic currents from a single neuron's dendrites add up to and may cause the neuron to generate an action potential. This neuron then synapses on other neurons, and so on. EEG represents summed postsynaptic potentials of neurons firing a rate of milliseconds. The recording of the brain's electrical activity obtained by using electrodes is also referred to as EEG (electro means electrical, encephelo means brain, and gram means record).

Recording of EEG

EEG is recorded by placing scalp electrodes on the scalp. The location of scalp electrodes used in recording EEGs are defined by an international agreement known as 10-20 system or International 10-20 system (see Figure 9). The technique is based on the relationship between the location of an electrode and the underlying area of the cerebral cortex. The numbers '10' and '20' refer to the fact that the distances between adjacent electrodes are either 10% or 20% of the total front-back or right-left distance of the skull. Each electrode site is identified by a letter to indicate the lobe and a number to indicate the hemisphere location. Table 6 shows each lobe and the electrode label. The letter 'C' is used for identification purposes only.

Table 6. Lobes and corresponding electrode label

Lobe	Electrode Label
Frontal	F
Temporal	T
Central	C
Parietal	P
Occipital	O

The label 'z' in Figure 9 refers to an electrode placed on the midline. The even numbers refer to electrode positions on the right hemisphere and the odd numbers refer to electrode positions on the left hemisphere.

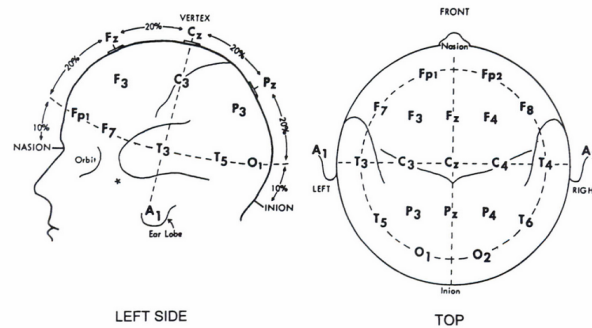


Figure 9. International 10-20 Electrode Placement System (Stern, 2005).

The scalp electrodes can be connected in different ways to the EEG monitoring system (Stern, 2005).

Characterization of EEG Signals

Graphically, an EEG signal is represented as a graph of the time varying voltage difference (see Figure 10) between an active electrode attached to the scalp and a reference electrode. The peak-to-peak amplitude of the scalp-recorded EEG signal in a healthy adult is under 100 microvolts (Gevins & Smith, 2006).

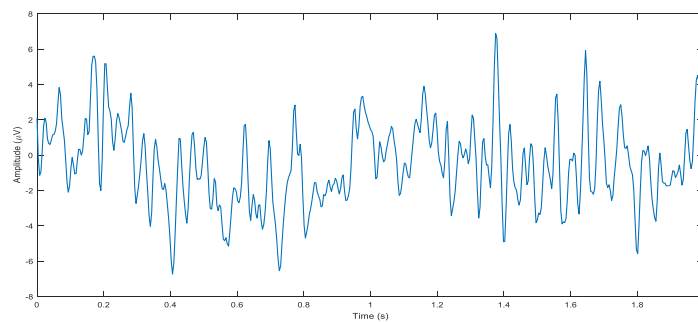


Figure 10. Time domain EEG signal.

EEG is usually described in terms of its repetitive rhythm (frequency). The rhythmic activity is divided into frequency ranges (bands). Every EEG signal is made up of five major brain waves each distinguished by its frequency range (see Figure 10). The brain waves are delta (0.5 - 4 Hz), theta (4 - 8 Hz), alpha (8-13Hz), beta (13- 30 Hz), and gamma (> 30Hz) (Siuly & Zhang, 2017). These bands are able to provide insight into a person's cognitive states such as attention and engagement. For example, the theta band has been associated with cognitive processes such as consciousness awareness (Klimesch et al., 2001), processing of new (episodic) information (Klimesch, 2012) and mental calculation (Fernández et al., 1995). The alpha band is associated with information processing (Klimesch, 2012), and mental effort (Klimesch, 1997). Alpha band power

decreases with increasing mental effort (Klimesch, 1997). Borghini, Astolfi, Vecchiato, Mattia, and Babiloni (2014) found an increase in theta band power and a decrease in alpha band power occurred in high mental workload. The gamma band is associated with various cognitive and motor functions.

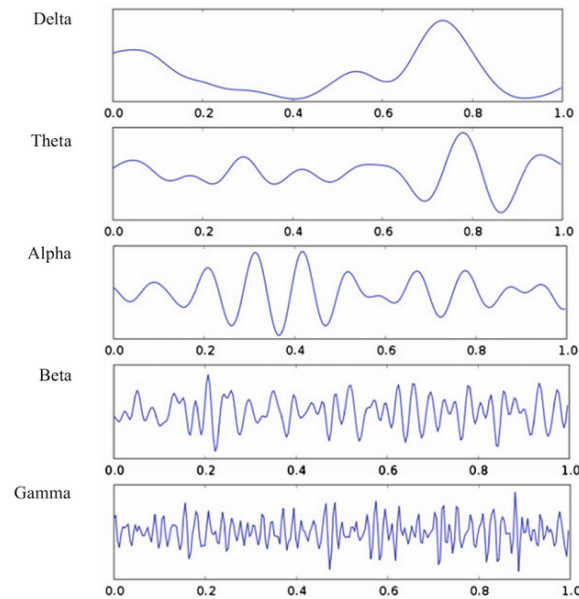


Figure 10. Main EEG waves (Siuly & Zhang, 2017).

EEG recordings may be affected by artifacts. Artifacts are signals that contaminate the EEG. They are not of cerebral origin, and are divided into physiological and non-physiological artifacts. Physiological artifacts are generated from the participant himself or herself and include eye and muscle movements. Non-physiological artifacts are generated from the participant's surroundings and include computers, power line, mobile phones and other electrical and electronic equipment (Sethi, Sethi, Torgovnick & Arsura, 2007). Preprocessing is required to remove artifacts, improve the signal-to-noise ratio, and prepare the signals in a suitable form for further processing.

Approaches for analyzing EEG data

Two main approaches exist for analyzing EEG data (van Vugt, Simen, Nystrom, Holmes & Cohen, 2012). One approach is to time lock EEG changes to sensory, motor or cognitive events. Event-related potentials (ERPs) are obtained by averaging these time-locked EEG signals. The EEG signals are averaged on the assumption that neural activities that are not time-locked to the sensory, motor or cognitive event sums to zero (Parasuraman & Rizzo, 2008). ERPs can be analyzed in time and frequency domains (Picton et al., 2000). ERP components are referred to by a letter and a number (see Figure 11 and Table 7). The letter indicates (N or P) indicates polarity (negative or positive), and the number indicates either the latency in milliseconds or the component's ordinal position in the waveform. (Nidal & Malik, 2014).

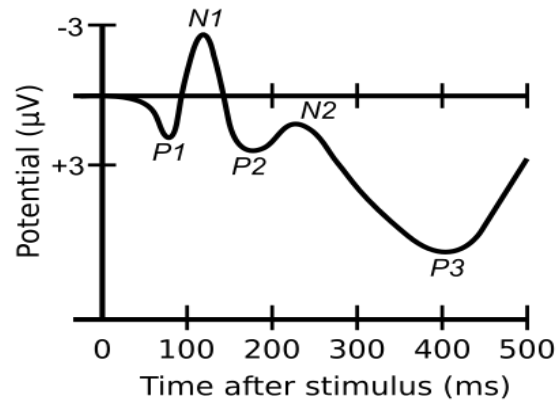


Figure 11. A waveform showing several ERP components (Event-related potential, n.d.).

For instance, N1 or N100 (indicating its latency is 100 ms after the stimulus and that it is negative) is a negative-going peak that often occurs about 100 milliseconds after a stimulus is presented. The P300 component of an ERP wave form is directly related to decision making (San Martín, 2012). It usually emerges in the late period after stimuli onset (300–600ms). In task relevant paradigms, the amplitude of P300 is generally considered as a representation of memory load (Karis, Fabiani, & Donchin, 1984; Azizian & Polich, 2007; Polich, 2007). The amplitude of P300 rises after the conclusion of a task related to stimuli differentiation reflecting attention reorganization. P300 is also related to working memory updating and learning processes. It has been suggested that this potential is elicited at the end of a cognitive episode in which the decision is made as to whether the stimulus is important or not. P300 latency may reflect the discrimination time needed to interpret a stimulus as important or unimportant (Vandoolaeghe, 1998). P300 is of particular interest in the present study. Less demanding tasks have relatively high P300 amplitude and relatively short peak latency. Demanding tasks reduce P300 amplitude and lengthen peak latency (Polich, 2007).

The other approach for analyzing EEG data is analyzing the spectral composition of EEG signals (Parasuraman & Rizzo, 2008; van Vugt et al., 2012). Spectral analysis involves quantifying amplitude or power of a certain frequency band. This approach is able to capture brain activity related to stimulus processing that is not time-locked to stimulus onset.

EEG/ERP correlates of decision making

Wang, Zheng, Huang, and Sun (2015) investigated the temporal electrophysiological changes and difference at the early stimuli assessment stage of decision making between the ambiguous and risky conditions. They designed a monetary gambling game in which the probabilities of outcomes were known (risk) or unknown (ambiguity) while the outcomes were varied across trials but balanced in pairs. They hypothesized that P300 would elicit smaller amplitude under ambiguity than under risk decision making. They recorded electroencephalography (EEG), and event related potential (ERP) technique applied to the experimental data. Their findings suggested that the P300 amplitude can be applied to reflect information processing at the early stage (the stage of assessment) of decision making.

Table 7. ERP components, their onset, topography, and functionality (adapted from Johnson & Proctor, 2013).

Component	Onset (ms)	Topography	Functionality
<i>Early processing</i>			
C1	65 – 90	Striate	Visual processing
P1	80 – 120	Extrastriate	Visual processing + attention
<i>Stimulus-processing related</i>			
P2	150 – 275	Centro-front parieto-occipital	Comparison with internal representation
N2c	200 – 300	Posterior	Degree of attention allocated to stimulus
N2pc	200 – 300	Posterior-contralateral	Attention allocation
<i>Stimulus-categorization</i>			
P300/P3b	300 – 600	Parietal	Stimulus evaluation and categorization
P3a	250 – 280	Frontocentral	Attention engagement, processing of novelty
N400	250 – 500	Centro-parietal	Semantic processing
P600	500 – 1200	Centro-parietal	Syntactic processing
<i>Response related</i>			
CNV	260 – 470	Vertex	Contingency between two stimuli
LRP	260 – 470	Centro-contralateral	Response preparation
ERN/Ne	80 – 150	Frontocentral	Error processing

Marzi, Righi, Ottonello, Cincotta, and Viggiano, (2014) assessed the temporal sequence of ERPs elicited by two different social evaluation tasks based on facial appearance and first impressions. EEG activity was recorded while participants performed a direct trustworthiness evaluation of briefly presented faces that varied along the trustworthiness dimension and a political election task ('would you vote for this political candidate?'), in which attention was not drawn on the concept of trust. Their findings showed that the N170 was not modulated by the emotional content of the faces but was influenced by the task performed. Furthermore, they found that the enhanced positivity for untrustworthy faces, not only started early, but persisted throughout the processing sequence, such that the amplitude of the EPN, 220–400 ms post-stimulus, and subsequent LPP (around 500 ms post-stimulus) were increased. They concluded that emotional responses, triggered by face evaluation, appear both quite early and at longer latencies and differ as a function of the evaluated dimension (trust or vote). The temporal profile of the ERP responses suggests that the evaluation of trust influences visual processing quite early and unfolds largely independently from structural encoding.

Cui, Chen, Wang, Shum, and Chan (2013) examined the neural correlates of uncertain decision making by administering an adapted Iowa Gambling Task (IGT) to participants. The adapted IGT allowed for the analysis of the choice evaluation, response selection, and feedback evaluation stages of uncertain decision making. EEG data were recorded. They used ERP to examine correlates of different stages of uncertain decision making. The obtained results were consistent with the notion that a positive somatic state was presented in the left hemisphere, and a negative somatic state was represented in the

right hemisphere. There were also anticipatory ERP effects that guided the participants' responses and provided evidence for the somatic hypothesis with more precise timing.

Portella et al. (2012) analyzed the possible correlation between early and late stages of information processing by assessing the latency to, and amplitude of, early and late ERP components, including P200, N200, premotor potential and P300, in healthy participants in the context of a visual oddball paradigm. They proposed that if the secondary processing of visual input occurs faster, then discrimination and classification process of the input, motor response processing, reorganization of attention and working memory update, and reaction time will occur faster. Their results demonstrated moderate positive correlations among the latency of P200, N200, motor response processing, and P300.

van Vugt et al. (2012) suggested that theta oscillations could have a fundamental role in perceptual decision making and specifically in the accumulation of evidence. They used a data-driven approach to find oscillatory correlates of evidence accumulation. This was based on the Drift Diffusion Model (DDM) which posits that to make a decision, a person accumulates information until it reaches a threshold, at which time they make the response that corresponds to that threshold. They recorded participants' EEG and developed a General Linear Model (GLM) method, in which they correlated the predicted accumulator dynamics with the EEG time series. They showed that EEG oscillations exhibit dynamics consistent with evidence accumulation in a perceptual decision making task. Accumulation dynamics were captured best by 4–9 Hz theta oscillations in a set of superior parietal channels, and they covaried with individual differences in DDM parameters fitted to behavioral data.

Davis, Wu, and Everhart (2011) sought to understand whether unique patterns of brain function in complex decision making (CDM) are detectable, and whether it is different from what has been observed with other types of decision making. They explored the spatial distribution of EEG responses during decision making across several simulated real-world contexts in discrete EEG frequency bands, with a particular emphasis on changes in the alpha bands. They hypothesized that they would observe unique decreases in alpha with CDM, and corresponding increases in other bands. Since CDM is thought to be highly dependent on frontal lobes, they also hypothesized that they would observe relatively greater activity at frontal electrode sites than at posterior sites. They recorded EEG from 30 scalp sites, during a series of baseline, eyes open fixation tasks, and CDM tasks in different contexts. They used five simulated real-world contexts. The CDM task was found to produce significant negative changes in all of the alpha bands. Significant correlations between EEG power spectra in different bands and at specific electrode sites and CDM task performance were observed. Also, significant negative changes occurred across alpha bands from the baseline to the CDM task in accordance with what they hypothesized. Decreases in alpha power are consistent with greater brain activity and cognitive task demand. Their findings indicate that it is possible to observe significant changes in EEG with CDM.

EEG Index of Hemispheric Asymmetry (Laterality Index)

According to Hugdahl and Westerhausen (2010) it is impossible to advance a theoretical understanding of brain-behavior relations without taking hemispheric asymmetry into account. There are two cerebral hemispheres (left hemisphere and right

hemisphere), and they exhibit different activation patterns when they are provoked by specific tasks (Brancucci, 2010). The left hemisphere is analytic, logical and learns by rules and is programmed instruction. The right hemisphere is intuitive, perceives wholes, and is specialized for organization and recognition of faces, patterns, and visuo-spatial data. The right hemisphere is responsible for emotional content than the left hemisphere (Cooksey 1996; Perlaki & Barchas, 2015).

When a HO is presented a task, one of his or her brain's hemispheres is usually more responsive to the task, and therefore gets relatively more activated than the other hemisphere. Higher activation reflects suppression of alpha activity. Thus, a decrease in alpha activity is observed on the hemisphere which is more involved in the task (Perlaki & Barchas, 2015). The laterality index may be computed from data in the frequency domain (Cajochen, Di Biase & Imai, 2008). After computing the total spectral energy for all EEG sub-bands is computed, alpha band percentage is given by the spectral energy in alpha band divided by total spectral energy for all sub-bands. The laterality index is given by Equation 16 (Perlaki & Barchas, 2015)

$$Laterality\ Index = \frac{(R - B_R) - (L - B_L)}{|(R - B_R) + (L - B_L)|} \quad (16)$$

where R is percentage of right hemisphere alpha during treatment, B_R is percentage of right hemisphere alpha during baseline, L is percentage of left hemisphere alpha during treatment, and B_L is percentage of left hemisphere alpha during baseline. A negative Laterality Index suggests more right hemisphere involvement. A positive Laterality Index suggests left hemisphere involvement (Perlaki & Barchas, 2015).

EEG Task Load Index (TLI)

EEG spectra have been found to respond to changes in working memory load (Lorenz & Parasuraman, 2003). EEG is able to provide a valid and objective index for mental effort (Serman & Mann, 1995). Berka et al. (2007) argued that EEG is the only physiological signal that accurately reflects subtle shifts in workload. Hogervorst et al. (2014) compared different physiological variables (EEG, skin conductance, respiration, ECG, pupil size and eye blinks) and found that EEG provided the best information about momentary workload of an individual. Alpha activity reflects different levels of workload. High alpha has been reported for low levels of workload (Fink, Grabner, Neuper, & Neubauer, 2005; Brouwer et al., 2012). The theta band has also been associated with workload or mental effort (Klimesch, 1997; Klimesch, 1999). The task load index (TLI), given by the ratio of frontal mid-line theta to parietal alpha, increases with mental workload (Gevins, Smith, McEvoy & Yu, 1997; Fairclough & Ewing, 2017).

$$TLI = \frac{Frontal\ Midline\ Theta\ Power}{Parietal\ Alpha\ Power} \quad (17)$$

Previous researchers have employed the TLI to evaluate workload. Vidulich, Stratton, Crabtree, and Wilson (1994) analyzed recorded EEG of participants in simulated air-to-ground-missions. In their experiment, they manipulated the type of information display. For the task condition that required the most cognitive load, they

found lower levels of alpha activity and higher levels of theta activity. Hankins and Wilson (1998) designed an experiment to collect cardiac, eye, EEG and subjective data during an actual flight scenario. The tasks required different piloting skills at several levels of mental workload. They found increased power in the theta band during an actual flight scenario in those flight segments, which required in-flight mental calculations. Recently, Kamzanova, Kustubayeva, and Matthews (2014) compared five EEG indices of alertness and workload as concomitants of vigilance decrement in performance. The EEG indices used in their study were alpha-1 suppression, alpha-2 suppression, engagement index, frontal theta, and task load index. They found lower-frequency alpha (8 to 10.9 Hz) and TLI to be most EEG indices sensitive to the vigilance task parameters. Frontal theta and EI were influenced by task workload only in the final period of the vigilance task. For the tasks used in their research, they concluded that lower-frequency alpha is the optimal index for monitoring vigilance

Table 8. Confusion matrix for a 2-class problem.

		Predicted Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

Classification of EEG

The purpose of classification is to assign objects to one of several predefined groups or classes. A classifier employs a learning algorithm to build a model that correctly predicts class labels of previously unknown data. Generally, a training dataset consisting of data with known labels are used to build a classification model. A test dataset with unknown class labels is applied to the classification model. The classifier's performance is evaluated based on the counts of test records correctly and incorrectly predicted by the classification model. These counts are tabulated in a table called confusion matrix. Table 8 shows the confusion matrix of a binary classification problem. TP, TN, FP, FN denote TP is the number of true positives, the number of true negatives, the number of false positives, and the number of false negatives respectively.

Several single value performance metrics may be derived from the confusion matrix. These include accuracy, error rate, sensitivity (true positive rate or recall), specificity (true negative rate), precision (positive predictive value), false positive rate, and F-score. Table 9 below describes each performance metric, and the corresponding equation.

Table 9. Classifier Performance Metrics.

Performance Metric	Description	Equation	
Accuracy	Answers the question: Overall, how often is the classifier correct?	$\frac{TP+TN}{TP+TN+FP+FN}$	(18)
Error Rate (Misclassification Rate)	Answers the question: Overall, how often is the classifier wrong?	$\frac{FP+FN}{TP+TN+FP+FN}$	(19)
Sensitivity (True Positive Rate or Recall)	Answers the question: When it's actually yes, how often does the classifier predict yes? Best sensitivity is 1.0, worst is 0.0	$\frac{TP}{TP+FN}$	(20)
Specificity (True Negative Rate)	Best specificity is 1.0, worst is 0.0 Answers the question: When it's actually no, how often does the classifier predict no?	$\frac{TN}{FP+TN}$	(21)
Precision (positive predictive value)	Best precision is 1.0, worst is 0.0 Answers the question: When it predicts yes, how often is the classifier correct?	$\frac{TP}{TP+FP}$	(22)
False positive rate	Best false positive rate is 0.0, the worst is 1.0 Can be calculated as 1 – specificity Answers the question: When it's actually no, how often does the classifier predict yes?	$\frac{FP}{TN+FP}$	(23)
F-score	Harmonic mean of precision and recall β is commonly 0.5, 1, or 2	$\frac{(1+\beta^2)(Precision \times Recall)}{(\beta^2 \times Precision + Recall)}$	(24)

Another performance metric is the Kappa statistic (Cohen, 1960), given by Equation 25. Unlike the other metrics, this metric takes into account random chance. It compares an observed accuracy with an expected accuracy. The metric is used to measure the reliability of two raters. In binary classification, one rater reflects the ground truth, and the other rater reflects the classifier used to perform the classification.

$$Kappa = \frac{(Observed\ Accuracy - Expected\ Accuracy)}{(1 - Expected\ Accuracy)} \quad (25)$$

Observed accuracy is the number of instances the classifier agreed with the ground truth divided by the total number of predictions (see Equation 26).

$$Observed\ Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \quad (26)$$

Expected accuracy is the accuracy any random classifier would be expected to achieve based on the confusion matrix (see Equation 27). It is found by multiplying the marginal frequency of a particular class for one rater by the marginal frequency for that same class for the second rater, and dividing by the total number of prediction. The marginal frequency for a certain class by a certain rater is the sum of all instances the rater indicated were that class. For example, (TP + FN) were labeled as Positive by the ground truth, and (TP + FP) were labeled as Positive by the classifier. In the same vein,

(FP + TN) were labeled as Negative by the ground truth, and (FN + TN) were labeled as Negative by the classifier.

$$Expected\ Accuracy = \frac{(TP + FN)(TP + FP)}{(TP + TN + FP + FN)} + \frac{(FP + TN)(FN + TN)}{(TP + TN + FP + FN)} \quad (27)$$

A kappa of 0.80 means there is 80% better agreement than chance. There is no universally accepted interpretation of the Kappa statistic. Table 10 shows kappa values and their interpretation considered by Landis and Koch (1977).

Table 10. Kappa values and the meaning (adapted from Landis & Koch,1977).

Kappa Value	Interpretation
> 0	No agreement
0.0 – 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 1	Almost perfect agreement

Classifiers using EEG features as inputs are able to discriminate between cognitive states with a significant level of accuracy. Classification algorithms that have been used to classify EEG include linear discriminant analysis (LDA), k-Nearest Neighbor, Artificial Neural Networks (ANNs), and SVM.

Linear Discriminant Analysis (LDA)

The LDA uses hyperplanes to separate data representing different groups or classes (Fukunaga, 2013). For a two-class problem, the side of the hyperplane on which the feature vector is determines its class (see Figure 12).

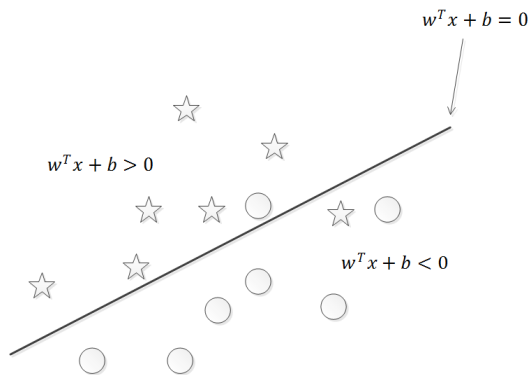


Figure 12. Linear classifier.

LDA assumes the data is normally distributed, and that the covariance matrices for both classes are equal. It seeks a projection that maximizes class separability, such that the distance between the two classes is maximized and interclass variance minimized (Lotte, Congedo, Lécuyer, Lamarche & Arnaldi, 2007; Fukunaga, 2013).

Mathematically, the hyperplane is represented by Equation 28.

$$w^T x + b = 0 \quad (28)$$

where w is the projection vector, x is the input feature vector and b is a bias. Let $x_1, \dots, x_n \in \mathbb{R}^d$ be n feature vectors. Input feature vectors are classified based on $\text{sign}(w^T x + b)$. The projection vector is given by Equation 29 (Blankertz, Lemm, Treder, Haufe & Müller, 2011).

$$w = \hat{\Sigma}_c^{-1}(\hat{\mu}_2 - \hat{\mu}_1) \quad (29)$$

where μ_i is the estimated mean of the class i , $\hat{\Sigma}_i$ is the average of the class-wise empirical covariance matrix, $\hat{\Sigma}_c$ is the estimated common covariance matrix

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i, \quad \hat{\Sigma}_i = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \hat{\mu})(x_i - \hat{\mu})^T \text{ and } \hat{\Sigma}_c = \frac{(\hat{\Sigma}_1 + \hat{\Sigma}_2)}{2} \quad (30)$$

Previous research (Garrett, Peterson, Anderson & Thaut, 2003; Bostanov, 2004; Bhardwaj, Gupta, Jain, Rani & Yadav, 2015) has used LDA to successfully classify EEG. For example, Bhardwaj et al. (2015) recorded EEG signals as emotions of participants were evoked by showing them pictures. LDA was used to successfully classify emotions into seven classes (happy, sad, anger, disgust, neutral, fear and surprised). LDA, however, has a drawback. It assumes normal distribution of the data, with equal covariance matrices for the classes. This assumption does not hold for EEG data.

Artificial Neural Network (ANN)

ANNs mimic the human brain. They are made up of interconnected nodes and directed links. The simplest ANN model is a perceptron (see Figure 13). Each input node is connected via a weighted link to the output node. The weighted link is used to emulate the strength of synaptic connection between neurons. The input nodes transmit the value they receive to the outgoing links without performing any transformations. The output node computes the weighted sum of the inputs, subtracts the bias term, and then produces an output that depends on the sign of the resulting sum.

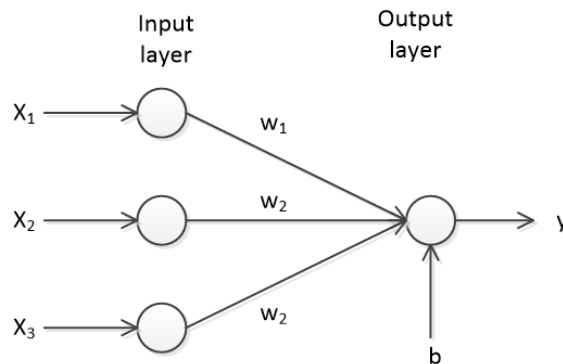


Figure 13. Example of a Perceptron.

A perceptron computes its output value, $\hat{y}$, by performing a weighted sum of its inputs, subtracting a bias factor b from the sum, and then examining the sign of the result. The output of a perceptron can be expressed mathematically by Equation 30.

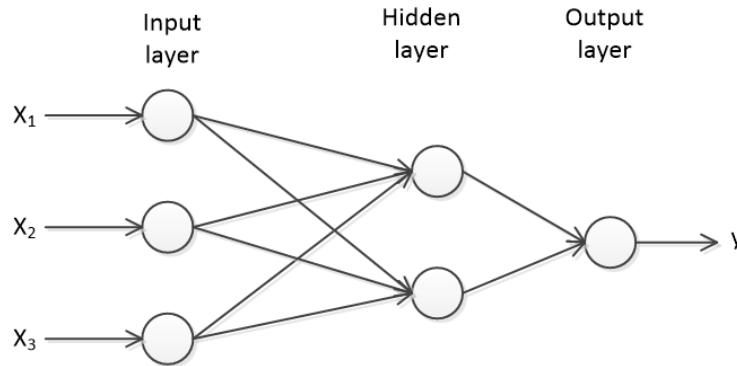
$$\hat{y} = \text{sign}(w_n x_n + w_n x_n + \dots + w_n x_n + w_0 x_0) = \text{sign}(\mathbf{w} \cdot \mathbf{x}) \quad (31)$$

where $w_0 = -b$, $x_0 = 1$, and $\mathbf{w} \cdot \mathbf{x}$ is the dot product between the weight vector $\mathbf{w}$ and the input vector $\mathbf{x}$.

During the training phase, the weight parameters $\mathbf{w}$ are adjusted until the outputs of the perceptron become consistent with the true outputs of training samples. The weight update equation is given by Equation 32.

$$w_j^{(k+1)} = w_j^{(k)} + \lambda(y_i - \hat{y}_i^{(k)})x_{ij} \quad (32)$$

where $w^{(k)}$ is the weight parameter associated with the i th input link after the k th iteration, λ is known as the learning rate, and x_{ij} is the value of the j th attribute of the training example x_i . The new weight $w^{(k+1)}$ is a combination of the old weight $w^{(k)}$ and a term proportional to the prediction error, $(y - \hat{y})$. If the prediction is correct, the weight is unchanged, else it is modified. The decision boundary of a perceptron is a linear hyperplane that separates the data into two classes. Thus, the perceptron algorithm is guaranteed an optimal solution for linearly separable classification problems. The algorithm fails to converge for nonlinearly separable data.



Example 14. Example of a multiple feed-forward ANN.

An ANN contains several intermediary layers, called hidden layers, between its input and output layers. In a feed-forward neural network, the nodes in one layer are connected only to the nodes in the next layer (see Figure 14). In a recurrent neural network, the links may nodes within the same layer or nodes from one layer to previous layers. ANN uses activation functions instead of the sign function. These include linear, sigmoid, and hyperbolic tangent functions. These functions allow the hidden and output nodes to produce outputs that are nonlinear in their input parameters. Thus the multilayer ANN is able to model complex relationships between the input and output variables. A more efficient algorithm that converges to the right solution is needed to learn the

weights of an ANN. The ANN algorithm determines a set of weights $\mathbf{w}$ that minimizes the total sum of squared errors.

Previous research (King, Nguyen, & Lal, 2006; Khare, Santhosh & Anand, 2008; Chai, Tran, Craig, Ling & Nguyen, 2014; Belakhdar, Kaaniche, Djmel & Ouni, 2016) has used ANN to successfully classify EEG. For example, Belakhdar et al. (2016) detected the occurrence of driver drowsiness onset based on ANN and using only one EEG channel. They conducted an experiment on ten human participants using nine features computed from one EEG channel using the Fast Fourier Transform (FFT). After introducing these features in an ANN classifier, [18] obtained a classification accuracy rate of 86.1% and 84.3% of drowsiness and alertness detection.

Support Vector Machines (SVM)

The principle of SVM classification is to construct an optimal hyperplane as the decision surface to identify the different classes, so as to maximize the spacing between them. They were originally designed for binary classification by finding the best hyperplane that separates all data points of one class from those of the other class. The best hyperplane for an SVM is the hyperplane with the largest margin between the two classes. The support vectors are the data points that are closest to the separating hyperplane.

Generally, an SVM maps an input vector $\vec{X}$ into a high dimensional feature space Z through a mapping function Φ . A hyperplane separates data belonging to different classes in Z . This hyperplane corresponds to a decision function containing Φ , a bias b , and a weighting vector $\vec{w}$. The SVM computes the best set of $\Phi, b, \vec{w}$ that generates a hyperplane minimizing classification errors based on a set of training data (Hornik, Meyer & Karatzoglou, 2006; Vapnik, 2000).

Linearly separable data ("hard-margin" linear SVM)

Given training data $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N \in R^n$, $y_1, y_2, \dots, y_N \in \{-1, +1\}$, there exists an infinite number of hyperplanes to separate negative instances from positive ones. SVMs find the hyperplane that maximizes the gap between data points on the boundaries. These data points are called support vectors (Figure 15).

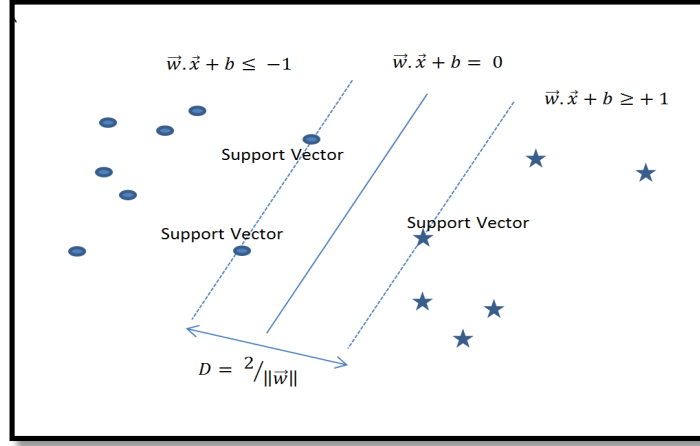


Figure 15. Hard-margin SVM.

In order to maximize the margin, D , there is the need to minimize $\frac{1}{2} \|\vec{w}\|$. Furthermore, to ensure that all instances are correctly classified,

$$\vec{w} \cdot \vec{x}_i + b \leq -1 \quad \text{if } y_i = -1 \quad (32)$$

$$\vec{w} \cdot \vec{x}_i + b \geq +1 \quad \text{if } y_i = +1 \quad (33)$$

These are equivalent to

$$y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1 \quad (34)$$

Thus given an instance x , the classifier is

$$f(\vec{x}) = \text{sign}(\vec{w} \cdot \vec{x} + b) \quad (35)$$

The primal formulation of the SVM optimization problem is given by

$$\text{Minimize } \frac{1}{2} \sum_{i=1}^n w_i^2 \quad \text{subject to } y_i(\vec{w} \cdot \vec{x}_i + b) - 1 \geq 0 \text{ for } i = 1, \dots, N \quad (36)$$

This is a convex quadratic programming (QP) optimization problem with n variables ($w_i, i = 1, \dots, n$), where n is the number of features in the dataset.

Non - linearly separable data ("soft-margin" linear SVM)

If the data is not linearly separable, an approach is to assign a "slack variable", ξ_i , to each instance, which can be thought of as distance from the separating hyperplane if an instance is misclassified and 0 otherwise (Figure 16).

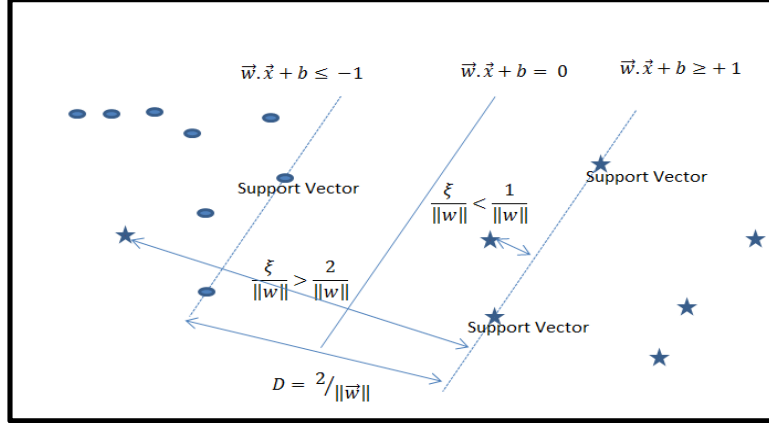


Figure 16. Soft-margin SVM.

For $0 < \xi \leq 1$ point is between margin and the correct side of the hyperplane. This is a margin violation. For $\xi > 1$ point is misclassified. The primal formula of the optimization problem becomes

$$\text{Minimize } \frac{1}{2} \sum_{i=1}^n w_i^2 + C \sum_{i=1}^N \xi_i \text{ subject to } y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1 - \xi_i \text{ for } i = 1, \dots, N \quad (37)$$

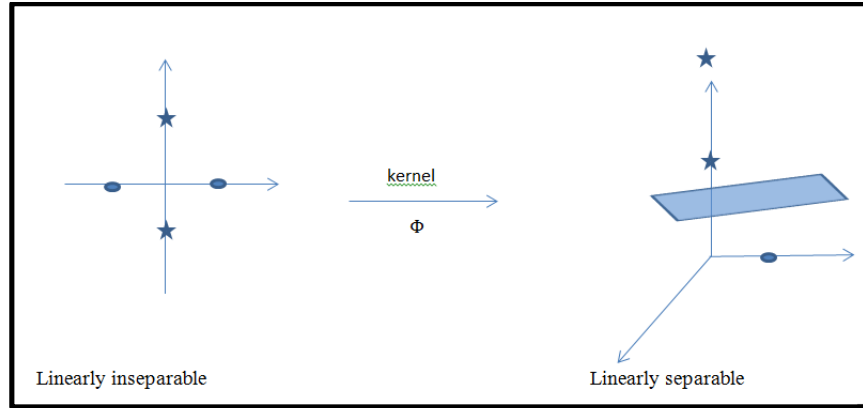


Figure 17. Non-linearly mapping of input data to a high-dimensional space.

Kernels are used to non-linearly map input data to a high-dimensional space (Figure 17). The new mapping is linearly separable. For the original data in the input space, $\vec{x}$, $f(\vec{x}) = \text{sign}(\vec{w} \cdot \vec{x} + b)$, and $\vec{w} = \sum_{i=1}^N \alpha_i y_i \vec{x}_i$. When the data is transformed in a higher dimensional feature space, $\Phi(\vec{x})$, $f(\vec{x}) = \text{sign}(\vec{w} \cdot \Phi(\vec{x}) + b)$, and $\vec{w} = \sum_{i=1}^N \alpha_i y_i \Phi(\vec{x}_i)$

$$f(\vec{x}) = \text{sign} \left(\sum_{i=1}^N \alpha_i y_i \Phi(\vec{x}_i) \cdot \Phi(\vec{x}) + b \right) \quad (38)$$

$$f(\vec{x}) = \text{sign} \left(\sum_{i=1}^N \alpha_i y_i K(\vec{x}_i, \vec{x}) + b \right) \quad (39)$$

The kernel function is given by

$$K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i)\Phi(\vec{x}_j) \quad (40)$$

is a dot product in some feature space. Four basic kernels used in SVM research are the linear kernel, the polynomial kernel, the radial basis function (RBF) kernel, and the sigmoid kernel. The RBF kernel given by

$$K(\vec{x}_i, \vec{x}_j) = e^{-\gamma\|\vec{x}_i - \vec{x}_j\|^2} \quad (41)$$

is generally chosen because (1) it can handle nonlinear relations between class labels and attributes, and (2) it has relatively small number of parameters.

Previous researchers have employed various feature extraction methods and the SVM algorithm for classification of EEG signals in different domains, and have obtained pretty good results. SVM has proved to be the superior classifier in many classification studies. For example, Bhardwaj et al. (2015) compared SVM and Linear Discriminant Analysis (LDA) classification of emotions into seven classes namely (happy, sad, anger, disgust, neutral, fear and surprised). Emotions were evoked in participants by showing them pictures and EEG signals corresponding to those events were recorded. With a 4-fold cross-validation, the SVM and LDA achieved an average overall accuracy of 74.13% and 66.50% respectively. Furthermore, Schlögl, Lee, Bischof and Pfurtscheller (2005) compared the performance of four classifiers (minimum distance analysis, LDA, k-nearest neighbor and SVM) applied to four class-EEG data. They used the adaptive autoregressive (AAR) process to model the EEG signal, and evaluated each classifier by Cohen's kappa coefficient. They found SVM as the most successful classifier. In the traffic safety domain, for example, Lee, Lee and Chung (2014) developed a driver monitoring system that used the most descriptive features selected by mutual information as input features for a Gaussian kernel SVM classifier. Using the wavelet packet transform (WPT) to extract each EEG frequency band from the wavelet coefficients, their SVM classifier achieved a classification accuracy of 98.6%. Wang, Jung, and Lin (2015) investigated drivers' focus of attention during dual task conditions (lane-keeping task and mathematical problem-solving task) based on recorded EEG data. They used SVM based on the radial basis function (RBF) kernel for classification. Classification accuracies of $84.6 \pm 5.8\%$ and $86.2 \pm 5.4\%$ were achieved on the math versus driving tasks, respectively.

Neuroergonomics

Neuroergonomics, defined as the study of brain and behavior (Parasuraman & Rizzo, 2007), allows researchers to study and develop new frameworks about humans and work than an approach based solely on the measurement of subjective perceptions or overt performance of the HO (Parasuraman & Rizzo, 2007). It asserts that the human brain must be examined during interaction with the environment to enable a complete understanding of cognition, action, and environment. Brain imaging techniques that used in neuroergonomics may be divided into two classes: those that measure cerebral hemodynamics, and those that measure electromagnetic activity (Parasuraman & Wilson,

2008). Examples of neuroimaging techniques that measure cerebral hemodynamics are functional magnetic resonance imaging (fMRI) and transcranial Doppler (TCD) sonography. Neuroimaging techniques that measure electromagnetic activity include EEG, event-related potentials (ERPs), and magnetoencephalography (MEG). Three criteria considered in determining the merits and demerits of each neuroimaging technique are spatial resolution, temporal resolution, and ease of use in human factors and ergonomics research. EEG scores highly in the last two criteria (Parasuraman & Wilson, 2008).

The ability to unobtrusively and continuously monitor operator mental states in a given task in an operational environment could be beneficial in finding more efficient and effective methods for humans to interact with technology (Berka et al., 2007). Task engagement, attention, and workload are examples of mental states that may be monitored. The information obtained can be exploited in an offline analysis for improving industrial production environments, the user interface of cars and for many other applications (Blankertz et al., 2010). Neuroergonomics may be applied in a variety of domains. These include aviation, driving, brain-computer interfaces, and virtual reality (Parasuraman & Rizzo, 2007).

EEG frontal theta and alpha power are modulated by variations in cognitive demand and effort. Ewing, Fairclough and Gilleade (2016) developed an adaptive game system that utilized real-time EEG to maximize player engagement in accordance with the level of game demand. In the first of two studies, they demonstrated the sensitivity of frontal theta and parietal alpha power to changing game demand levels. In the second study they evaluated the performance of an adaptive game with respect to system behavior and user experience. Their research provides a model for developing a biocybernetic loop.

Giraudet, Imbert, Bérenger, Tremblay & Causse (2015) used evaluated human machine interface (HMI) design in air traffic control. They assumed that HMI design will impact an operator's ability to process auditory stimuli because vision and hearing share attentional resources. Participants performed a simulated ATC task as the main task and responded to the occurrence of tones as the secondary task. They compared behavioral and neural responses to two different HMI designs. Their results showed that an enhanced HMI design releases attentional resources of an operator, enabling him or her to efficiently process other additional critical stimuli.

Schubert et al. (2008) investigated visual attention and lapses in vigilance during driving. They designed an experiment to account for the complexity of driving, which requires attention to both visual and auditory stimuli. They recorded EEG as participants performed three different primary tasks: fixate at a cross (baseline), video presentation of a driving scene (passive driving), and active driving requiring the use of a steering wheel in a lane change test. The secondary task which was kept constant across all three tasks was for participants to respond to high or low auditory stimuli by pressing left or right buttons. They measured driving performance as the time needed to press the correct button. They found that reaction time increased gradually from the baseline to the active

driving task. Also, alpha and gamma power decreased gradually from the baseline to the active driving task. They suggested that for competing tasks, fast reactions to auditory stimuli can be performed only on the expense of a less efficient engagement of visual processing.

Wilson and Russell (2003) investigated whether ANN and Stepwise Discriminant Analysis (SWDA) could be used to produce accurate estimates of operator functional state using psychophysiological measures. The classifiers were able to discriminate among four levels of mental workload during a simulated air traffic control task. Both classifiers achieved very high levels of classification. They suggested adaptive aiding as the logical application of their system. Wilson and Russell (2007) continued their previous study. They implemented adaptive automation and collected psychophysiological data while participants performed an uninhabited aerial vehicle (UAV) task. They used ANN to detect high and low mental workload. Their results proved that in the context of a UAV task, a workload classification system can be integrated into a successful adaptive aiding system.

CHAPTER 3. HYPOTHESES DEVELOPMENT

In chapter 2, I provided an extensive review of literature on the Cognitive Continuum Theory (CCT), and the Cognitive Fit Theory (CFT). The CCT and CFT have similarities and differences. These two theories complement each other in many ways. Both theories emphasize how the combination of information presentation format and nature of the task influence cognitive strategy and ultimately judgement performance. According to the CFT, cognitive strategy is selected by matching information displays to tasks. When there is a match between the format of data presentation and task type, HOs are able to form the right mental representations and use cognitive strategies to fit the presented data. This fit minimizes the HO's cognitive effort leading to effective and efficient performance (Vessey, 1994). The CCT, on the other hand, posits that task characteristics (surface and depth) induce certain types of cognitive strategy. Correspondence between task characteristics and cognitive strategy leads to better judgment accuracy (Hammond et al., 1987). Furthermore, both theories posit that performance is influenced by the level of correspondence between information presentation and cognitive strategies. One difference between the two theories is the classification of cognitive strategy. Whereas the CCT assumes two types of cognitive strategies (perceptual and analytical), the CCT assumes three cognitive strategies (intuitive, analytical, and quasirational).

The CFT is a suitable theoretical framework for understanding the relationship between information displays and task types (Basole et al., 2016), while the CCT is a coalescing theory for the field of human judgment and decision making (Cooksey, 1996). To the best of my knowledge, no research has been carried out to explore the potential of using both theories in a single study. In the present study, I investigate the effect of surface and depth characteristics on cognitive mode and judgment performance (see Table 11). Cognitive mode (intuition or analysis) is not directly observable but can be inferred from properties (Hammond et al., 1987) that discriminate between intuition and analysis. Moreover, tasks possess certain characteristics that induce a cognitive mode (Hammond et al., 1987).

Table 11. Combination of CCT and CFT.

		Surface Task Characteristics (Presentation Format)	
		Graph (Intuition-inducing)	Table (Analysis-inducing)
Depth Task Characteristics (Task Type)	Intuition- inducing Task	1. Congruence or Cognitive Fit 2. Intuitive Cognition 3. Better Performance than no congruence	1. No Congruence or Cognitive Fit 2. Quasirationality Cognition 3. Lower Performance than congruence
	Analysis- inducing Task	1. No Congruence or Cognitive Fit 2. Quasirationality Cognition 3. Lower Performance than congruence	1. Congruence or Cognitive Fit 2. Analysis Cognition 3. Better Performance than no congruence

In the present study the surface task characteristics (presentation format) are graph and table. The graph is intended to induce intuition cognition, and the table is intended to induce analysis cognition. I will use two types of tasks: intuition-inducing, and analysis inducing. The intuition-inducing task has six cues, and is presented in a relatively shorter time (Hammond et al., 1987). The analysis-inducing task has three cues, and is presented in a relatively longer time (Hammond et al., 1987).

Cognitive properties make it possible to differentiate intuition from analysis. In the present study, I use three properties (consistency, response time, and error distribution) to differentiate between the modes of cognition. High consistency is characteristic of analysis cognition. Faster response rate is characteristic of intuition cognition. A leptokurtic error distribution is reflective of analysis cognition (Dunwoody et al., 2000). Judgment performance is computed from achievement, r_a , in the lens model equation. I will measure perceived workload with the NASA Task Load Index (Hart & Staveland, 1988). Based on these measurements, I make the first five hypotheses:

H1: Graphical information display will induce intuition cognition and numerical information display will induce analytic cognition.

H2: The dominant cognitive mode is analysis when both surface and depth task characteristics have analysis inducing properties.

H3: The dominant cognitive mode is intuition when both surface and depth task characteristics have intuition inducing properties.

H4: Judgmental performance is positively related to the correspondence between task properties and cognitive mode.

H5: Cognitive load is lower when there is correspondence between surface and depth task properties.

The inferences and generalization made thus far, are from behavioral (Lens Model Parameters and Response Time), and subjective (NASA TLX) measures. Behavioral measures alone may not produce much information on the cognitive state of the HO (Parasuraman, 2003; Frey et al., 2013). Physiological measures, along with behavioral and subjective measures can be used to determine whether a stimulus produces the same human responses as compared to a baseline (Kramer & Weber, 2000).

To the best of our knowledge, no study has combined judgment analysis techniques with neurophysiological measures to understand human-computer interaction. I seek to bridge this gap in the present study. Furthermore, the effects of intuition-inducing and analysis-inducing information displays on judgment and decision making performance, and corresponding neural correlates have not been explored. In the present study, EEG will be recorded during judgment and decision making task execution.

Intuition cognition emphasizes right brain hemisphere, whereas analysis cognition emphasizes left brain hemisphere (Cooksey, 1996; Perlaki & Barchas, 2015). The TLI increases with mental workload (Gevins, Smith, McEvoy & Yu, 1997). P300 is of particular interest in the present study. Less demanding tasks have relatively high P300 amplitude and relatively short peak latency. Demanding tasks reduce P300 amplitude and lengthen peak latency (Polich, 2007). Based on these, I make the following three hypotheses:

H6: The left brain hemisphere is more activated during analysis inducing task, and the right brain is more activated during intuition inducing task.

H7: TLI is lower when there is correspondence between surface and depth task properties.

H8: P300 amplitude is relatively high and peak latency is relatively short for intuition-inducing tasks than for analysis-inducing tasks.

Lastly, classifiers using EEG features as inputs are able to discriminate between cognitive states with a significant level of accuracy. SVMs will be employed to classify EEG signals recorded during execution of the intuition- and analysis-inducing tasks

CHAPTER 4. METHODOLOGY

This chapter describes the methodology that will be used in the present study. Judgment analysis techniques with physiological measures will be used to explore effects of task type and data presentation format on cognitive mode and judgment performance. Judgment analysis emphasizes representative experimental design. Representative experimental design adopts an idiographic-nomothetic approach. This approach first examines each participant's judgment, statistically test it before generalizing trends. Design considerations in judgment analysis include the availability of an ecological criterion, whether or not feedback is utilized, the experience of the judges, cue distributions and cue intercorrelations, and the number of cue profiles to judge versus number of cues to make available (Cooksey, 1996).

In the present study, I use the double system Lens Model. No cognitive feedback is utilized. The judges are unfamiliar and inexperienced, and will need to be trained.

Description of Judgment Task

Advances in communication and sensor technology have accelerated the capability to remotely monitor water quality. Real time monitoring of water quality provides decision makers vital early warning information to enable them to respond appropriately (Glasgow, Burkholder, Reed, Lewitus & Kleinman, 2004). Online Water Quality Monitoring (OWQM) involves collecting real-time water quality data from monitoring sites in a source water or drinking water distribution system in order to detect unusual water quality. OWQM has two applications: Source Water Monitoring (SWM), and Distribution System Monitoring (DSM).

Analysis Inducing Task used in this Study

SWM involves monitoring water quality parameters upstream in source water or watersheds. Water quality parameters that can be monitored with online instruments include dissolved oxygen (DO), Fecal Coliform (FC), pH, Biochemical oxygen demand (BOD), Temperature Change (TC), Total phosphate, Nitrates, Turbidity, and Total Solids (Berlemann, 2013; Sun, Xia, Xu, Guo & Sun, 2016; Heddum, 2017).

Water quality index (WQI), a unitless number that ranges from 1 to 100 (Hallock, 2002), may be used to determine source water quality (Berlemann, 2013). It is based on nine water quality parameters (see Table 12). Two variables are required to compute WQI: Q value, and Weight Factor. The Q value for each parameter ranges between 0 and 100, and may be derived from a table or curve. Each parameter's data collected from an online sensor is converted to a corresponding Q value. The weighting factor allows for one parameter to have more impact on the overall water quality score than other factors. The WQI is computed by summing the products of the Q-value and weight factor of each parameter (see Equation 42)

$$WQI = \sum_{i=1}^n Q_i W_i \quad (42)$$

where Q_i is the Q value of parameter i , W_i is the weight of parameter i , and n is the number of parameters. Higher values of WQI indicate better water quality. Table 13 below explains WQI ranges and their meaning.

Table 12. Water quality parameters, description, and weights (adapted from Berlemann, 2013).

Water Quality Parameter	Description	Weight Factor
Dissolved oxygen	Measure of the amount of life-sustaining oxygen dissolved in the water Absence of dissolved oxygen indicates pollution	0.17
Fecal coliform	Fecal coliform is a form of bacteria found in human and animal waste A high count suggests pathogenic organisms are present	0.16
pH	Measure of the of acid–base balance in the water pH changes can indicate increased amounts of pollutants	0.11
Biochemical oxygen demand	Measure of the amount of food for bacteria that is found in water. High values are indicative of less desirable water quality	0.11
Temperature change	Physical, biological, and chemical characteristics of a river are directly affected by temperature	0.10
Total phosphate	Measure of the total amount of phosphate potentially available in a given water supply. Increase can increase amount of algae in water	0.10
Nitrates	Increase can increase amount of algae in water Nitrates can be harmful to humans.	0.10
Turbidity	Measure of the dispersion of light in a column of water due to suspended matter The higher the turbidity, the cloudier the water appears.	0.08
Total solids	Measure of the solid materials dissolved in the river water Too high or too low concentrations may limit growth and lead to the death of many aquatic life forms	0.07

Table 13. WQI range and meaning (adapted from Berlemann, 2013).

WQI Range	Water Quality
0-25	Very bad
25-50	Bad
50-70	Fair
70-90	Good
90-100	Excellent

In a situation where parameters cannot be analyzed, a new WQI is computed by distributing the weight factors of these parameters over the remaining parameters (Berlemann, 2013; Srivastava & Kumar, 2013). The new WQI is computed with Equation 43.

$$WQI = \sum_{a=1}^m \left(\frac{Q_a W_a}{\sum W_a} \right) \quad (43)$$

where Q_a is the Q value of an available parameter a , W_a is the weight factor of an available parameter a , and m is the number of available parameters.

Computation of WQI is inherently analytical. In the present study, I assume only three (DO, pH, and Turbidity) of the nine parameters are available. The weights of the three parameters are shown in Table 14. The new WQI is given by Equation 44.

Table 14. Three available parameters and their weight factors.

Parameter	Weight Factor
DO	$\frac{0.17}{(0.17 + 0.11 + 0.08)} \approx 0.50$
pH	$\frac{0.11}{(0.17 + 0.11 + 0.08)} \approx 0.30$
Turbidity	$\frac{0.08}{(0.17 + 0.11 + 0.08)} \approx 0.20$

$$WQI = 0.50Q_{DO} + 0.30Q_{pH} + 0.20Q_{Turbidity} \quad (44)$$

Values of Q_{DO} , Q_{pH} , and $Q_{Turbidity}$ will be randomly assigned from the values in Table 15. The product of each parameter's weight factor and Q value will be computed, and rounded to the nearest whole number. Participants will be asked to sum up three products to determine WQI. WQIs below 50 will be judged as bad quality, and those above 50 will be judged as good quality. The data will be presented in both tabular and graphical formats (see Figures 18a and 18b).

For example, for say $Q_{DO} = 56$, $Q_{pH} = 24$, $Q_{Turbidity} = 34$, the values that will be presented to the participant 28 (the product of 0.50 and 56), 7 (the product of 0.30 and 24 rounded to the nearest whole number), and 6 (the product of 0.20 and 34 rounded to the nearest whole number). The participant will have to add these three numbers (28, 7, and 6) up to arrive at the WQI. In this example, the value of WQI is 41.

DO	pH	Turbidity
28	7	6

Figure 18a. Analysis-inducing task presented in table format

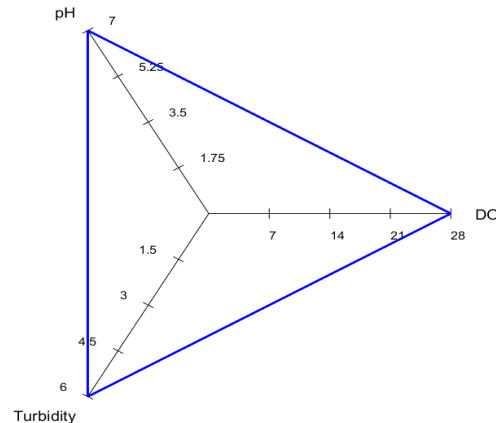


Figure 18b. Analysis-inducing task presented in graph format

Table 15. National Sanitation Foundation Q values (adapted from (Berlemann, 2013)).

Dissolved Oxygen	Q_{DO}	pH	Q_{pH}	Turbidity	$Q_{Turbidity}$
0	0	<2.0	0	0	97
10	8	2.0	2	5	84
20	13	3.0	4	10	76
30	20	4.0	8	>15	70
40	30	5.0	24	15	68
50	43	6.0	55	20	62
60	56	7.0	90	25	57
70	77	7.2	92	30	53
80	88	7.5	93	35	48
85	92	7.7	90	40	45
90	95	8.0	82	50	39
95	97.5	8.5	67	60	34
100	99	9.0	47	70	38
105	98	10.0	19	80	25
110	95	11.0	7	90	22
120	90	12.0	2	100	17
130	85	>12.0	0	>100	5
140	78				
>140	50				

Inducing-inducing Task used in this Study

Drinking water utilities manage water quality in drinking water distribution systems. They monitor baseline conditions in real-time. These utilities use OWQM data to detect unusual water quality conditions in their distribution systems. Water quality parameters that they monitor include chlorine, fluoride, nitrate, pH, turbidity, and total dissolved solids (TDS). The US Environmental Protection Agency (EPA) has water quality standards that must be met by drinking water utilities. Some contaminants and their maximum contaminant levels are shown in Table 16 below. HOs visualizing the data must ensure no contaminant exceeds its maximum contaminant level (MCL).

This task is inherently intuitive by nature. HOs predict water quality by comparing water quality parameters with their baseline values. In the present study we will present all six cues to participants in both graphical and tabular formats (see Figures 19a and 19b). Water quality will be judged as bad when any of the six parameters exceeds its MCL, otherwise, water quality will be judged as good.

Table 16. Contaminants and the maximum contaminant levels (Table of Regulated Drinking Water Contaminants, 2016).

Contaminant	Maximum Contaminant Level
Chloride	250 mg/L
Fluoride	2.0 mg/L
pH	6.5 – 8.5
Total Dissolves Solids	500 mg/L
Nitrate	10 mg/L
Turbidity	0.1 – 1 NTU

Parameter	Value	Maximum Contaminant Level
Chloride	200	250
Fluoride	1	2
pH	7	6.5 – 8.5
Total Dissolved Solids	400	500
Nitrate	8	10
Turbidity	0.1	1

Figure 19a. Intuitive-inducing task presented in table format

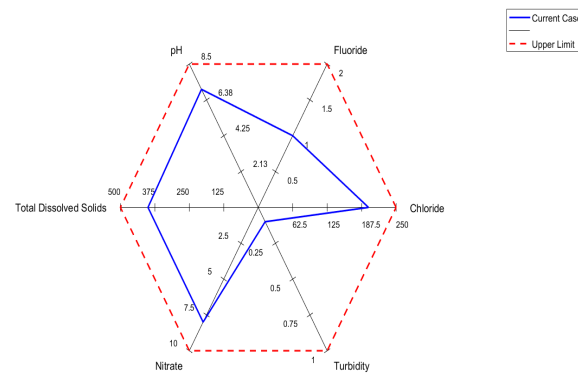


Figure 19b. Intuitive-inducing task presented in graph format

Participants

Participants (graduate and undergraduate students) from the North Carolina A&T State University will participate in the study. Each participant will be required to read and sign an informed consent form. All participants will have no history of neuropsychiatric disorders, and will not be taking psychoactive medications.

Experimental Design

All participants will undergo two conditions (graph and table) but will be assigned into two separate groups (intuition-inducing task or analysis-inducing task). The between-subjects factor is task type (intuition-inducing task or analysis-inducing task), and the within-subjects factor is presentation format (graph and table). Table 17 below shows the experimental design. The aim of this design is to understand if there is an interaction between task type and presentation format on cognitive mode, and ultimately judgment performance. Other dependent variables are response time, EEG recorded as participants perform the judgment tasks, and subjective NASA TLX rating administered after each task.

It will be counterbalanced in the order of block (stimuli) presentation in order to reduce the bias that could result from the order in which the treatments are provided.

At the end of the experiment, data will be analyzed using mixed ANOVA provided model assumptions are not violated, or if violated can be corrected. However, when violations cannot be corrected, I will use nonparametric tests (e.g. Wilcoxon signed-rank test) to analyze the data.

Table 17. Experimental design.

		Surface Task Characteristic (Presentation Format)	
		Table	Graph
Depth Task Characteristic (Task Type)	Intuition-inducing Task	(Intuition-inducing task in table format)	(Intuition-inducing task in graph format)
	Analysis-inducing Task	(Analysis-inducing task in table format)	(Analysis-inducing task in graph format)

Materials

Participants will be fitted with g.GAMMAcap (g.tec, Austria) which uses the International 10-20 system for the EEG electrode placement locations on the scalp. The electrodes will be connected and amplified with a g.HIamp (g.tec, Austria) amplifier. The stimuli will be presented in the Presentation software (Neurobehavioral Systems, www.neurobs.com/presentation) and EEG signals will be recorded and visually inspected with g.Recorder (g.tec, Austria). NASA Task Load Index questionnaire will be used to collect subjective workload data.

Procedure

Participants will first complete an informed consent form and demographic questionnaire. They will be given a brief introduction to the tasks. Participants will be told to base their judgments of water quality on the information provided. Each cue will be described but no indication of its importance will be given.

The outcome feedback will be used to train the participants. Training will consist of 2 blocks of each level of presentation format. Each block will contain thirty trials. Participants will not be fitted EEG equipment during the training session.

After training, participants will be fitted with the EEG equipment. For each group of participants, there will be three treatments: baseline (eye-opened resting condition), judgment task presented in graph format, and judgment task presented in table format. In the baseline task, participants will be instructed to relax, and fixate a blank screen for one minute. Apart from the baseline, the other two treatments will each contain four blocks, with twenty trials in each block. To avoid order effects, the block of trials and the trials in a block will be presented in pseudorandom order. To help encourage participants to remain engaged in the task throughout the experiment, short blocks of trials will be used. Participants will be instructed to click on the left mouse button for “bad water quality” judgments, and the right mouse button for “good water quality” judgments.

Each trial will begin with a cross fixation for 1 second, after which the stimulus will be presented. For the analysis-inducing task condition, presentation time of stimuli in each trial will be 120 s (see Figures 20a and 20b). A trial ends when participants click

the mouse button before or at the stipulated duration of 120s. A cross fixation of variable duration (2~4s) is presented between trials.

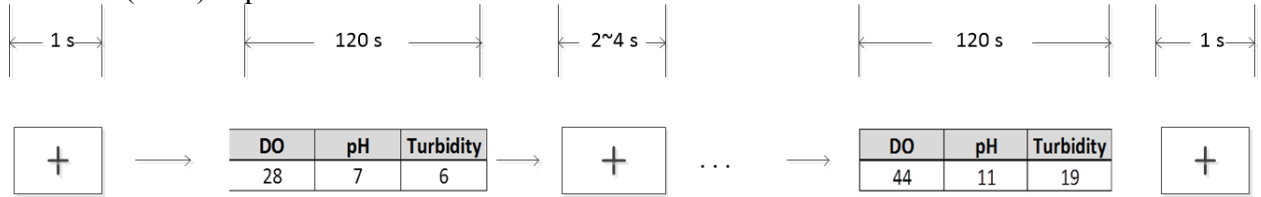


Figure 20a. Trials in analysis-inducing task condition. Data is presented in table format.

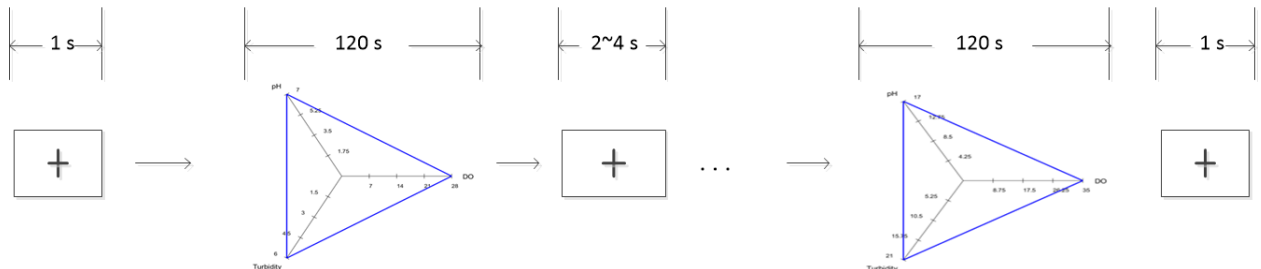


Figure 20 b. Trials in analysis-inducing task condition. Data is presented in graph format.

For the intuition-inducing task condition, presentation time of stimuli in each trial will be 60 s (see Figures 21a and 21b). A trial ends when participants click the mouse button before or at the stipulated duration of 60 s. A cross fixation of variable duration (2~4s) is presented between trials.

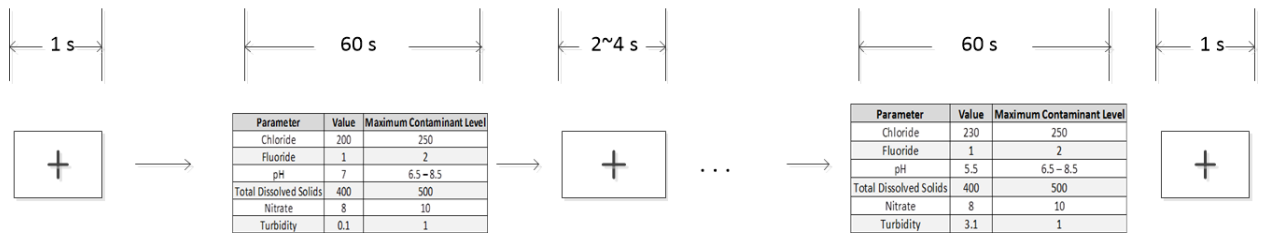


Figure 21a. Trials in intuition-inducing task condition. Data is presented in table format.

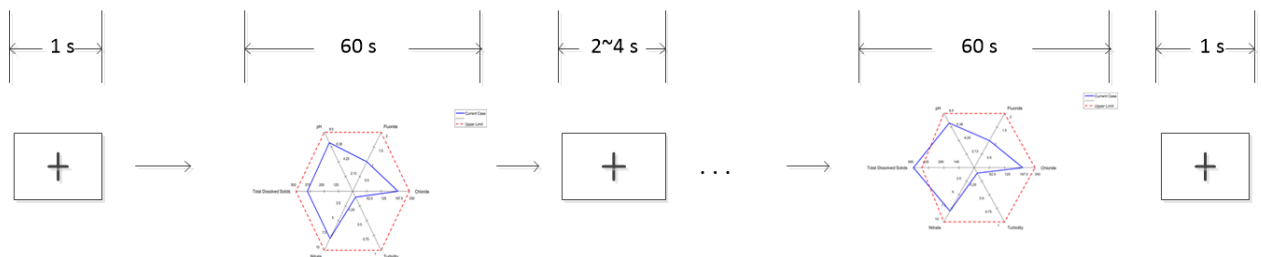


Figure 22b. Trials in intuition-inducing task condition. Data is presented in graph format.

References

- Anderson, J. R. (2013) "The Architecture of Cognition". Mahwah, New Jersey: Psychology Press.
- Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., ... & Babiloni, F. (2016). Adaptive automation triggered by EEG-based mental workload index: a passive brain-computer interface application in realistic air traffic control environment. *Frontiers in human neuroscience*, 10.
- Azizian, A., & Polich, J. (2007). Evidence for attentional gradient in the serial position memory curve from event-related potentials. *Journal of cognitive neuroscience*, 19(12), 2071-2081.
- Bacic, D. (2014). The Role of Cognitive Effort in Decision Performance Using Data Representations;; a Cognitive Fit Perspective. Doctor of Business Administration, Cleveland State University, Cleveland.
- Basole, R. C., Huhtamäki, J., Still, K., & Russell, M. G. (2016). Visual decision support for business ecosystem analysis. *Expert Systems with Applications*, 65, 271-282.
- Belakhdar, I., Kaaniche, W., Djmel, R., & Ouni, B. (2016, March). Detecting driver drowsiness based on single electroencephalography channel. In 2016 13th International Multi-Conference on Systems, Signals & Devices (SSD) (pp. 16-21). IEEE.
- Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., ... & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, 78(5), B231-B244.
- Berlemann, A. (2013). Using a water quality index to determine and compare creek water quality. *Journal (American Water Works Association)*, 105(6), E291-E298.
- Bhardwaj, A., Gupta, A., Jain, P., Rani, A., & Yadav, J. (2015, February). Classification of human emotions from EEG signals using SVM and LDA Classifiers. In *Signal Processing and Integrated Networks (SPIN)*, 2015 2nd International Conference on (pp. 180-185). IEEE.
- Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K. R. (2011). Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage*, 56(2), 814-825.
- Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., ... & Mueller, K. R. (2010). The Berlin brain-computer interface: non-medical uses of BCI technology. *Frontiers in neuroscience*, 4, 198.
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, 44, 58-75.
- Bostanov, V. (2004). BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. *IEEE Transactions on Biomedical engineering*, 51(6), 1057-1061.
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological review*, 62(3), 193.
- Brancucci, A. (2010). Electroencephalographic and magnetoencephalographic indices of hemispheric asymmetry. *The two Halves of the Brain: information processing in the cerebral hemispheres*. The MIT Press, Cambridge, 211-251.

- Brouwer, A. M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural engineering*, 9(4), 045008.
- Cajochen, C., Di Biase, R., & Imai, M. (2008). Interhemispheric EEG asymmetries during unilateral bright-light exposure and subsequent sleep in humans. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 294(3), R1053-R1060.
- Cheng, S. Y., & Hsu, H. T. (2011). *Mental Fatigue Measurement Using EEG*. Rijeka: INTECH Open Access Publisher.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37-46.
- Cooksey, R. W. (2000). Commentary on cognitive adaptation and its consequences: a test of cognitive continuum theory'. *Journal of Behavioral Decision Making*, 13(1), 55.
- Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. Academic Press.
- Chai, R., Tran, Y., Craig, A., Ling, S. H., & Nguyen, H. T. (2014, August). Enhancing accuracy of mental fatigue classification using advanced computational intelligence in an electroencephalography system. In 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 1318-1341). IEEE.
- Christensen, J. C., Estep, J. R., Wilson, G. F., & Russell, C. A. (2012). The effects of day-to-day variability of physiological data on operator functional state classification. *NeuroImage*, 59(1), 57-63.
- Cui, J. F., Chen, Y. H., Wang, Y., Shum, D. H., & Chan, R. C. (2013). Neural correlates of uncertain decision making: ERP evidence from the Iowa Gambling Task. *Frontiers in human neuroscience*, 7.
- Davis, C. E., Hauf, J. D., Wu, D. Q., & Everhart, D. E. (2011). Brain function with complex decision making using electroencephalography. *International Journal of Psychophysiology*, 79(2), 175-18.
- Dunn, C., & Grabski, S. (2001). An Investigation of Localization as an Element of Cognitive Fit in Accounting Model Representations. *Decision Sciences*, 32(1), 55-94.
- Dunwoody, P. T., Haarbauer, E., Mahan, R. P., Marino, C., & Chu-Chun, T. (2000). Cognitive adaptation and its consequences: A test of cognitive continuum theory. *Journal of Behavioral decision making*, 13(1), 35.
- Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgement and choice. *Annual review of psychology*, 32(1), 53-88.
- Epstein, S. (2008). Intuition from the perspective of cognitive-experiential self-theory. *Intuition in judgment and decision making*, 23, 37.
- Estep, J. R., & Christensen, J. C. (2011, August). Physiological cognitive state assessment: Applications for designing effective human-machine systems. In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE (pp. 6538-6541). IEEE.
- Evans, J. S. B. (2012). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning*, 18(1), 5-31.
- Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3), 223-241.

Event-related potential. (n.d.). Retrieved February 06, 2017, from https://en.wikipedia.org/wiki/Event-related_potential

Ewing, K. C., Fairclough, S. H., & Gilleade, K. (2016). Evaluation of an Adaptive Game that Uses EEG Measures Validated during the Design Process as Inputs to a Biocybernetic Loop. *Frontiers in human neuroscience*, 10.

Fairclough, S. H., & Ewing, K. (2017). The effect of task demand and incentive on neurophysiological and cardiovascular markers of effort. *International Journal of Psychophysiology*.

Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a characterization of adaptive systems: A framework for researchers and system designers. *Human Factors*, 54(6), 1008-1024.

Fernández, T., Harmony, T., Rodríguez, M., Bernal, J., Silva, J., Reyes, A., & Marosi, E. (1995). EEG activation patterns during the performance of tasks involving different components of mental calculation. *Electroencephalography and clinical Neurophysiology*, 94(3), 175-182.

Fink, A., Grabner, R. H., Neuper, C., & Neubauer, A. C. (2005). EEG alpha band dissociation with increasing task demands. *Cognitive brain research*, 24(2), 252-259.

Frey, J., Mühl, C., Lotte, F., & Hachet, M. (2013). Review of the use of electroencephalography as an evaluation method for human-computer interaction. arXiv preprint arXiv:1311.2222.

Fu, S., & Parasuraman, R. (2006). Event-related potentials (ERPs) in neuroergonomics. OXFORD SERIES IN HUMAN-TECHNOLOGY INTERACTION, 32.

Fukunaga, K. (2013). Introduction to statistical pattern recognition. Academic press.

Garrett, D., Peterson, D. A., Anderson, C. W., & Thaut, M. H. (2003). Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Transactions on neural systems and rehabilitation engineering*, 11(2), 141-144.

Gevins, A., & Smith, M. E. (2006). Electroencephalography (EEG) in neuroergonomics. *Neuroergonomics: The brain at work*, 15-31.

Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cerebral cortex*, 7(4), 374-385.

Giraudet, L., Imbert, J. P., Bérenger, M., Tremblay, S., & Causse, M. (2015). The neuroergonomic evaluation of human machine interface design in air traffic control using behavioral and EEG/ERP measures. *Behavioural brain research*, 294, 246-253.

Glasgow, H. B., Burkholder, J. M., Reed, R. E., Lewitus, A. J., & Kleinman, J. E. (2004). Real-time remote monitoring of water quality: a review of current applications, and advancements in sensor, telemetry, and computing technologies. *Journal of Experimental Marine Biology and Ecology*, 300(1), 409-448.

Hallock, D., 2002. A Water Quality Index for Ecology's Stream Monitoring Program. Washington Department of Ecology. A Department of Ecology Report,02-03-052. <https://fortress.wa.gov/ecy/publications/summarypages/0203052.html> (accessed Feb. 3, 2017).

Hamm, R. M., & Yang, H. (2016). Alternative lens model equations for dichotomous judgments about dichotomous criteria. *Journal of Behavioral Decision Making*.

Hammond, K. R. (2006). Beyond rationality: The search for wisdom in a troubled time. Oxford University Press.

- Hammond, K. R. (1996). Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice. Oxford University Press.
- Hammond, K. R. (1980). The integration of research in judgment and decision theory (No. CRJP-226). COLORADO UNIV AT BOULDER CENTER FOR RESEARCH ON JUDGMENT AND POLICY.
- Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE transactions on Systems, Man, and Cybernetics*, 17(5), 753-77.
- Hammond, K. R., & Stewart, T. R. (1975). 10. SOCIAL JUDGMENT THEOR Y. *Human judgment and decision processes*, 271.
- Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, space, and environmental medicine*, 69(4), 360-367.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139-183.
- Heddam, S. (2017). Fuzzy Neural Network (EFuNN) for Modelling Dissolved Oxygen Concentration (DO). In *Intelligence Systems in Environmental Management: Theory and Applications* (pp. 231-253). Springer International Publishing.
- Hogarth, R. M. (2002). Deciding analytically or trusting your intuition? The advantages and disadvantages of analytic and intuitive thought. *The Advantages and Disadvantages of Analytic and Intuitive Thought* (October 2002). UPF Economics and Business Working Paper, (654).
- Hogervorst, M. A., Brouwer, A. M., & van Erp, J. B. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload.
- Holcomb, R. C. (2011). Improving Judgment Performance and Cognitive Systems Engineering by Examining the Relationship between Task Properties and Cognitive Mode. GEORGE MASON UNIVERSITY.
- Hornik, K., Meyer, D., & Karatzoglou, A. (2006). Support vector machines in R. *Journal of statistical software*, 15(9), 1-28.
- Hursch, C. J., Hammond, K. R., & Hursch, J. L. (1964). Some methodological considerations in multiple-cue probability studies. *Psychological review*, 71(1), 42.
- Hugdahl, K., & Westerhausen, R. (2010). *The two halves of the brain: information processing in the cerebral hemispheres*. MIT press.
- Johnson, A., & Proctor, R. (2013). *Neuroergonomics: A cognitive neuroscience approach to human factors and ergonomics*. Palgrave Macmillan.
- Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, 58(9), 697.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.
- Kamzanova, A. T., Kustubayeva, A. M., & Matthews, G. (2014). Use of EEG workload indices for diagnostic monitoring of vigilance decrement. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 56(6), 1136-1149.

- Karis, D., Fabiani, M., & Donchin, E. (1984). "P300" and memory: Individual differences in the von Restorff effect. *Cognitive Psychology*, 16(2), 177-216.
- Khare, V., Santhosh, J., & Anand, S. (2008, December). Performance comparison using five ANN methods for classification of EEG signals of two mental states. In 2008 Annual IEEE India Conference (Vol. 1, pp. 7-10). IEEE
- Kelton, A. S., Pennington, R. R., & Tuttle, B. M. (2010). The effects of information presentation format on judgment and decision making: A review of the information systems research. *Journal of Information Systems*, 24(2), 79-105.
- King, L. M., Nguyen, H. T., & Lal, S. K. L. (2006, August). Early driver fatigue detection from electroencephalography signals using artificial neural networks. In Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE (pp. 2187-2190). IEEE.
- Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in cognitive sciences*, 16(12), 606-617.
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, 29(2), 169-195.
- Klimesch, W. (1997). EEG-alpha rhythms and memory processes. *International Journal of Psychophysiology*, 26(1), 319-340.
- Klimesch, W., Doppelmayr, M., Yonelinas, A., Kroll, N. E., Lazzara, M., Roehm, D., & Gruber, W. (2001). Theta synchronization during episodic retrieval: neural correlates of conscious awareness. *Cognitive Brain Research*, 12(1), 33-38.
- Kramer, A. F., & Weber, T. (2000). Applications of psychophysiology to human factors. *Handbook of psychophysiology*, 2, 794-814.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.
- Lee, B. G., Lee, B. L., & Chung, W. Y. (2014). Mobile healthcare for automatic driving sleep-onset detection using wavelet-based EEG and respiration signals. *Sensors*, 14(10), 17915-17936.
- Lorenz, B., & Parasuraman, R. (2003). Human operator functional state in automated systems: the role of compensatory control strategies. *NATO SCIENCE SERIES SUB SERIES I LIFE AND BEHAVIOURAL SCIENCES*, 355, 224-237.
- Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of neural engineering*, 4(2), R1.
- Marzi, T., Righi, S., Ottonello, S., Cincotta, M., & Viggiano, M. P. (2014). Trust at first sight: evidence from ERPs. *Social cognitive and affective neuroscience*, 9(1), 63-72.
- Moshman, D. (2000). Diversity in reasoning and rationality: Metacognitive and developmental considerations. *Behavioral & Brain Sciences*, 23, 689-690.
- Nidal, K., & Malik, A. S. (Eds.). (2014). *EEG/ERP Analysis: Methods and Applications*. CRC Press.
- Parasuraman, R. (2003). Adaptive automation matched to human mental workload. *NATO Science Series Sub Series I Life and Behavioural Sciences*, 355, 177-193.

Parasuraman, R., de Visser, E., Clarke, E., McGarry, W. R., Hussey, E., Shaw, T., & Thompson, J. C. (2009). Detecting threat-related intentional actions of others: effects of image quality, response mode, and target cuing on vigilance. *Journal of Experimental Psychology: Applied*, 15(4), 275.

Parasuraman, R., & Rizzo, M. (2007). Introduction to neuroergonomics. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The brain at work* (pp. 3–12). New York: Oxford University Press.

Parasuraman, R., & Rizzo, M. (Eds.). (2008). *Neuroergonomics: The brain at work*. Oxford University Press.

Parasuraman, R., & Wilson, G. F. (2008). Putting the brain to work: Neuroergonomics past, present, and future. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3), 468-474.

Perlaki, K. M., & Barchas, P. (2015). *Processing of Information Acquired at a Preconscious Level of Awareness: Instruction and Sex Effects on Hemispheric Lateralization and Accuracy*.

Peters, J. R., Srivastava, V., Taylor, G. S., Surana, A., Eckstein, M. P., & Bullo, F. (2015). Human supervisory control of robotic teams: integrating cognitive modeling with engineering design. *IEEE Control Systems*, 35(6), 57-80.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., ... & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, 37(2), 127-152.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, 118(10), 2128-2148.

Portella, C., Machado, S., Arias-Carrión, O., Sack, A. T., Silva, J. G., Orsini, M., ... & Ribeiro, P. (2012). Relationship between early and late stages of information processing: an event-related potential study. *Neurology international*, 4(3).

Pritchett, A. R., & Bisantz, A. M. (2006). Measuring the fit between human judgments and alerting systems: A study of collision detection in aviation. *Adaptive Perspectives on Human-Technology Interaction: Methods and Models for Cognitive Engineering and Human-Computer Interaction*, 91-104.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic bulletin & review*, 11(6), 988-1010.

Rothrock, L., Koubek, R., Fuchs, F., Haas, M., & Salvendy, G. (2002). Review and reappraisal of adaptive interfaces: toward biologically inspired paradigms. *Theoretical Issues in Ergonomics Science*, 3(1), 47-84.

San Martín, R. (2012). Event-related potential studies of outcome processing and feedback-guided learning. *Frontiers in human neuroscience*, 6.

Scerbo, M. W., Freeman, F. G., & Mikulka, P. J. (2003). A brain-based system for adaptive automation. *Theoretical Issues in Ergonomics Science*, 4(1-2), 200-219.

Schlögl, A., Lee, F., Bischof, H., & Pfurtscheller, G. (2005). Characterization of four-class motor imagery EEG data for the BCI-competition 2005. *Journal of neural engineering*, 2(4), L14.

Schubert, R., Tangermann, M., Haufe, S., Sannelli, C., Simon, M., Schmidt, E. A., ... & Curio, G. (2008). Parieto-occipital alpha power indexes distraction during simulated car driving. *International journal of psychophysiology*, 69(3), 214.

- Sethi, N. K., Sethi, J., Torgovnick, E., & Arsura, E. (2007). Physiological and non-physiological EEG artifacts. *The Internet Journal of Neuromonitoring*, 5(1).
- Siuly, S., Li, Y., & Zhang, Y. (2017). *EEG Signal Analysis and Classification: Techniques and Applications*. Springer.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological bulletin*, 119(1), 3.
- Speier, C. (2006) "The influence of information presentation format on complex task decision-making performance". *International Journal of Human-Computer Studies*. Vol. 64. pp. 1115-1131.
- Srivastava, G., & Kumar, P. (2013). Water quality index with missing parameters. *International Journal of Research in Engineering and Technology*, 2(4), 609-614.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral & Brain Sciences*, 23, 645-665.
- Sterman, M. B., & Mann, C. A. (1995). Concepts and applications of EEG analysis in aviation performance evaluation. *Biological psychology*, 40(1), 115-130.
- Stern, J. M. (2005). *Atlas of EEG patterns*. Lippincott Williams & Wilkins.
- Stevenson, R. J. (1997). Deductive reasoning and the distinction between implicit and explicit processes. *Current Psychology of Cognition*, 16, 222-229.
- Stewart T: Notes on a form of the lens model equation for logistic regression analysis. Brunswik Society Meeting; 2004.
- Stewart, T. R. (2001). The lens model equation. *The essential Brunswik: Beginnings, explications, applications*, 357-362.
- Subha, D. P., Joseph, P. K., Acharya, R., & Lim, C. M. (2010). EEG signal analysis: a survey. *Journal of medical systems*, 34(2), 195-212.
- Sun, W., Xia, C., Xu, M., Guo, J., & Sun, G. (2016). Application of modified water quality indices as indicators to assess the spatial and temporal trends of water quality in the Dongjiang River. *Ecological Indicators*, 66, 306-312.
- Sun, H., Xiang, Y., Sun, Y., Zhu, H., & Zeng, J. (2010, October). On-line EEG classification for brain-computer interface based on CSP and SVM. In *Image and Signal Processing (CISP)*, 2010 3rd International Congress on (Vol. 9, pp. 4105-4108). IEEE.
- Table of Regulated Drinking Water Contaminants. (2016, October 04). Retrieved February 06, 2017, from <https://www.epa.gov/ground-water-and-drinking-water/table-regulated-drinking-water-contaminants>
- Ting, C. H., Mahfouf, M., Nassef, A., Linkens, D. A., Panoutsos, G., Nickel, P., ... & Hockey, G. R. J. (2010). Real-time adaptive automation system based on identification of operator functional state in simulated process control operations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 40(2), 251-262.
- Tucker, L. R. (1964). A suggested alternative formulation in the developments by Hirsch, Hammond, and Hirsch, and by Hammond, Hirsch, and Todd. *Psychological review*, 71(6), 528.

- Tuttle, B. M., & Kershaw, R. 1998. Information Presentation and Judgment Strategy from a Cognitive Fit Perspective. *Journal of Information Systems*, 12(1): 1.
- Umanath, N. S., & Vessey, I. (1994). Multiattribute Data Presentation and Human Judgment: A Cognitive Fit Perspective *Decision Sciences*, 25, 795-824.
- Vandoolaeghe, E., van Hunsel, F., Nuyten, D., & Maes, M. (1998). Auditory event related potentials in major depression: prolonged P300 latency and increased P200 amplitude. *Journal of affective disorders*, 48(2), 105-113.
- van Vugt, M. K., Simen, P., Nystrom, L. E., Holmes, P., & Cohen, J. D. (2012). EEG oscillations reveal neural correlates of evidence accumulation. *Frontiers in neuroscience*, 6.
- Vapnik, V. N. (2000). Methods of pattern recognition. In *The nature of statistical learning theory* (pp. 123-180). Springer New York.
- Vessey, I. (1991). Cognitive fit: A theory- based analysis of the graphs versus tables literature. *Decision Sciences*, 22(2), 219-240.
- Vessey, I. (1994) "The effect of information presentation on decision making: A cost-benefit analysis". *Information & Management Vol. 27*. pp. 103-119.
- Vessey, I. , & Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition. *Information Systems Research*, 2 (1), 63-84.
- Vidulich, M. A., Stratton, M., Crabtree, M., & Wilson, G. (1994). Performance-based and physiological measures of situational awareness. *Aviation, Space, and Environmental Medicine*.
- Wang, L., Zheng, J., Huang, S., & Sun, H. (2015). P300 and Decision Making under Risk and Ambiguity. *Computational Intelligence and Neuroscience*, 501, 108417.
- Wang, Y. K., Jung, T. P., & Lin, C. T. (2015). EEG-based attention tracking during distracted driving. *IEEE transactions on neural systems and rehabilitation engineering*, 23(6), 1085-1094.
- Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering psychology & human performance*. Psychology Press.
- Wilson, G. F., & Russell, C. A. (2003). Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(3), 381-389.
- Wilson, G. F., & Russell, C. A. (2007). Performance enhancement in an uninhabited air vehicle task using psychophysiological determined adaptive aiding. *Human Factors*, 49(6), 1005-1018.
- Wood, R. E. 1986. Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, 37(1):60-82.
- Zudilova-Seinstr, E. (2007). On the role of individual human abilities in the design of adaptive user interfaces for scientific problem solving environments. *Knowledge and Information Systems*, 13(2), 243-270.

STUDY 2–1

ELECTROENCEPHALOGRAPHY (EEG) CLASSIFICATION OF COGNITIVE TASKS BASED ON TASK ENGAGEMENT INDEX

Abstract — The application of autonomous systems is on an increase, and there is the need to optimize the fit between humans and these systems. While operators must be aware of the autonomous system's dynamic behaviors, the autonomous systems must in turn base their operations, among other things, on an on-going knowledge of operators' cognitive state, and its application domain. Psychophysiology allows for the use of physiological measurements to understand an operator's behavior by non-invasively recording peripheral and central physiological changes while the operator behaves under controlled conditions. Electroencephalography (EEG) is a psychophysiological technique for studying brain activation. In the present study, EEG task engagement index, defined as the ratio of beta to (alpha + theta), are used as inputs to an artificial neural network (ANN) to allow identification and classification of mental engagement. Six separate feedforward ANN with single hidden layer trained by backpropagation were designed to classify five mental tasks for each of six participants. The average classification accuracy across the six participants was 88.67 %. The results show that differences in cognitive task demand do elicit different degrees of mental engagement, which can be measured through the use of the task engagement index.

I. INTRODUCTION

The application of automated and autonomous systems is on an increase, and there is the need to optimize the fit between humans and these systems. For example, in an autonomous or highly automated work environment where the operator's role is primarily monitoring the system, there is very little overt operator performance to observe in order to determine whether or not the operator needs adaptive aiding. Adaptive aiding improves performance of system by providing automation when the operator needs it [1]. Furthermore, in such environments, the functional state of operators may be monitored in order to determine when to provide the aid. Previous research suggests that physiological measures such as EEG ([2],[3]) may be sensitive to unpredictable changes in cognitive load making them useful for adaptive automation.

EEG signals represent summed postsynaptic potentials of neurons firing a rate of milliseconds. Graphically, an EEG is a graph of the time varying voltage difference between an active electrode attached to the scalp and a reference electrode [4]. EEG may be analyzed in the time (event related potentials), frequency (spectral) or time–frequency domain. Event related potentials (ERPs) represent the brain's neural response to specific sensory, motor or cognitive events. They are computed by averaging EEG epochs time–locked to a specific event [5]. However, ERPs fail to capture brain activity related to stimulus processing that is not time–locked to event onset [6]. The frequency bands in an EEG signal are delta (less than 4Hz), theta (4– 8Hz), alpha (8 – 12Hz), beta (12– 30Hz), and gamma (30 – 80Hz) [6]. During highly controlled cognitive tasks, the spectral composition of EEG changes in response to changes in task difficulty or level of alertness [7]. Spectral analysis involves quantifying amplitude or power of a certain frequency

band. This approach is able to capture brain activity related to stimulus processing that is not time-locked to stimulus onset. Time-frequency analysis involves quantifying the spectral energy over time. It maps the 1-dimensional EEG time-domain signal into a 2-dimensional function of time and frequency.

In order to determine the operator functional state, physiological measures recorded during task performance are combined by a classifier. Classifiers such as artificial neural networks (ANN) have been used to obtain estimates of operator functional state. Data that represent the cognitive states of interest from each operator separately or data obtained from a group of similar operators are used to train the classifier. Afterwards, the trained classifier is continuously fed with physiological data during performance of a task. The classifier uses these data to estimate the operator functional state. When the classifier detects that the operator needs assistance, it notifies the system so that the operator will be provided the right automation [1].

Pope et al. [8] developed the first brain-based adaptive system that used an EEG-based task engagement index. The system used a bio-cybernetic loop that was formed by changing levels of automation in response to changes in mental workload demands. Of the candidate indices that they evaluated, Pope et al. [8] found that the engagement index based on the ratio $\beta/(\alpha + \theta)$ was the most sensitive. Further studies undertaken by Freeman et al. [9] confirmed the effectiveness of the $\beta/(\alpha + \theta)$ index. They reported improved performance in a vigilance task when the engagement index was used to drive changes in the stimulus presentation. In another study, Berka et al. [10] found that engagement and workload increased significantly during encoding of verbal and image-learning and memory test compared with recognition/recall period.

Other researchers have explored the task engagement index in determining the cognitive states of human participants. For example, in one study Szafrir and Mutlu [11] employed the task engagement index formula to study how participants interacted with an adaptive robotic agent that monitored attention in real time and adapted its behavior to improve the discourse. The agent monitored attention levels of participants and detected attention drops. In another study Szafrir and Mutlu [12] designed and implemented a system that employed the engagement index to monitor participants' attention to educational material in real time. This adaptive system suggested the optimal review topic. Wilson and Russell [3] presented an EEG-augmented reading system that monitored the engagement levels of children in real time, and provided contextual BCI training sessions to improve the children's reading engagement. McMahan et al. [14] compared three different EEG engagement indices (frontal θ , ratio of β to $\alpha + \theta$, ratio of frontal θ to parietal α) by assessing user engagement during various video game modalities. They found $\beta/(\alpha + \theta)$ to be the best algorithm for calculating the engagement levels of players playing video games.

Research has shown that classifiers using physiological features are able to determine the level of cognitive activity in tasks with a high level of accuracy. King et al. [15] employed ANN that used the magnified gradient function (MGF) technique to detect and classify early driver fatigue. The MGF reduced the time required for training by modifying the standard back propagation (SBP) algorithm. The ANN classified professional driver fatigue with 81.49% accuracy and non-professional driver fatigue with 83.06% accuracy. Khare et al. [16] compared the performance of five ANN methods

(Gradient Descent Back Propagation, Levenberg-Marquardt, Resilient Back Propagation, Conjugate Learning Gradient Back Propagation, and Gradient Descent Back Propagation with momentum) for classification of planning of right hand movement with respect to an awake relaxed state. Wavelet packet transform (WPT) was used for feature extraction of the relevant EEG signals. The Resilient Back Propagation method showed the highest performance with the accuracy of 95%. Chai et al. [17] examined the classification between fatigue and alert states using an autoregressive (AR) model-based power spectral density (PSD) as the features extraction method and fuzzy particle swarm optimization with cross mutated of artificial neural network (FPSOCM-ANN) as the classification method. Using 32-EEG channels, they obtained a classification accuracy of 80.51%. Classification using fewer (eleven frontal) EEG channels resulted in 75.65% classification accuracy. [18] detected the occurrence of driver drowsiness onset based on the Artificial Neuronal Network (ANN) and using only one EEG channel. They conducted an experiment on ten human participants using nine features computed from one EEG channel using the Fast Fourier Transform (FFT). After introducing these features in an ANN classifier, Belakhdar et al. [18] obtained a classification accuracy rate of 86.1% and 84.3% of drowsiness and alertness detection.

Although previous studies are suggestive of the potential of classifiers using physiological features to determine the level of cognitive activity in tasks, some questions still remain. Of particular interest is the question of whether or not task engagement indices (TEIs) can be used as features for classification of cognitive tasks. In particular, it has yet to be demonstrated that TEIs can serve as features to optimally discriminate recorded brain signals. We address this issue by employing TEIs as input to ANN to allow identification and classification of mental engagement. We propose that differences in task demand will elicit different degrees of mental engagement, which could be measured through the use of the task engagement index.

This study is outlined as follows. First, we describe the EEG data acquisition and preprocessing, time-frequency representation, and classification processes under the methodology section. Next, we present the results and discussion. Finally, we draw conclusions.

II. METHODOLOGY

A. EEG DATA ACQUISITION AND PRE-PROCESSING

The EEG data used in this study was originally collected by [19], and was downloaded from the website of the Brain-Computer Interfaces Laboratory of the Colorado State University [20]. The participants were seated in an Industrial Acoustics Company sound controlled booth with dim lighting and noise-less fan (for ventilation). An electro-cap elastic electrode cap was used to record EEG signals from positions C3, C4, P3, P4, O1, and O2 (shown in Figure 1), defined by the 10–20 system of electrode placement [21]. The electrodes were connected through a bank of Grass 7P511 amplifiers and bandpass filtered from 0.1–100Hz. Data was and recorded at a sampling rate of 250 Hz with a Lab Master 12 bit A/D converter mounted in an IBM-AT computer. Eye blinks were detected by means of a separate channel of data recorded from two electrodes placed above and below the participant's left eye. Continuous EEGs were recorded for 10s during each task.

In the present study, data from six participants performing five mental tasks were analyzed. The five tasks are baseline, multiplication, letter composition, geometric figure rotation, and visual counting. For each task, participants were instructed not to vocalize or gesture in any way. The baseline task was an eye opened task in which participants were asked to relax as much as possible. This task served as a reference task against which the other tasks were to be compared. In the multiplication task, participants were given nontrivial multiplication problems, such as 49 times 78, and were asked to solve them without vocalizing or making any other physical movements. In the letter composition task, participants were instructed to mentally compose a letter to a friend or relative without vocalizing. In the geometric figure rotation task, participants were asked to visualize a particular three-dimensional block figure being rotated about an axis. In the visual counting task, the participants were asked to imagine a blackboard and to visualize numbers being written on the board sequentially.

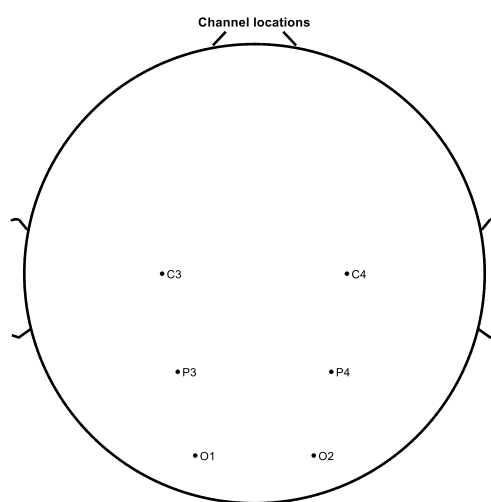


Fig. 1. Electrode placement.

Data was recorded for 10 seconds during each task and each task was repeated five times per session. With a 250 Hz sampling rate, each 10 second trial produced 2,500 samples per channel. Data from six participants were used. All data analyses were performed offline with MATLAB [22], EEGLAB [23], and custom code.

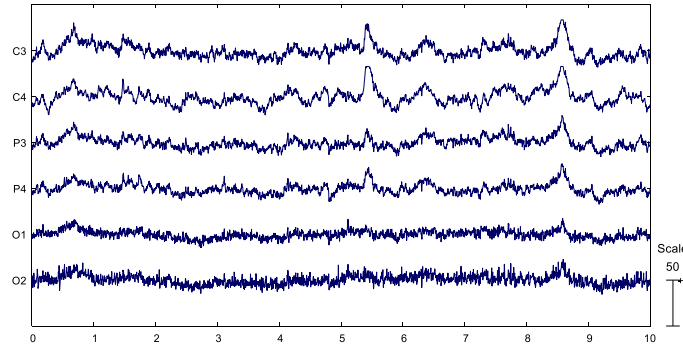


Fig. 2. Recorded raw EEG signal for participant 1 during the multiplication task.

Figure 2 shows the raw recorded EEG for participant 1 during the multiplication task. To remove linear trends, the data was high-pass filtered with basic Finite Impulse Response (FIR) filter with 1 Hz as the lower edge frequency. The data was re-referenced to average. The independent component analysis (ICA) can help correct EEG data contaminated by signals of non-neural origin. In the present study, the SASICA software [24] was used to reject artifact independent components (ICs) before EEG data analysis.

B. TIME-FREQUENCY REPRESENTATION

It is extremely important to find a suitable representation of EEG signals in order to reliably discriminate and understand extracted relationships [25]. The original recorded EEG data used in the present study were time domain signals.

Several techniques exist for spectral estimation and representation of EEG signals. Among these is the Fast Fourier Transform (FFT) which allows for the efficient estimation of the component frequencies in data from a discrete set of values sampled at a fixed rate. However, for signals whose frequencies change in time (like EEG), the FFT has disadvantages. The FFT cannot provide simultaneous time and frequency localization, which means that the power spectrum does not provide information about when certain frequencies occur in the signal. It is therefore not very useful for analyzing time-variant, nonstationary signals like EEG. This problem is overcome by using the Short Term Fourier Transform (STFT). The STFT segments the signal into narrow time intervals, and takes the Fourier Transform (FT) of each segment. Each FT provides the spectral information of a separate time-slice of the signal, providing simultaneous time and frequency information. The STFT was used in the present study. It was used to estimate the power spectrum of the EEG bands. This was implemented with the spectrogram function in MATLAB. The spectrogram computes an FFT-based spectral

estimate over each sliding window and allows for the visualization of how the frequency content of the signal changes over time. The spectrogram function divides a signal into segments. Long segments or windows (also known as narrowband spectrogram) provide better frequency resolution whereas short segments (also known as wideband spectrogram) provide better time resolution. In the present study, the data was segmented into quarter-second windows with a 50% overlap of the previous segment and 50% of the next. The Hann window which has good frequency resolution and reduced spectral leakage (Cerna & Harvey, 2000) was used. The log power, computed as $10 * \log_{10}(\text{power})$, of the data in each window was computed for the theta band (4-8Hz), alpha band (8-12Hz), and beta band (12-30Hz).

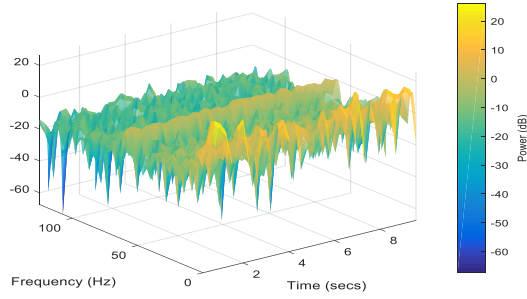


Fig. 3. 3-D spectrogram visualization of the multiplication task for participant 1.

Figure 3 shows the 3-D spectrogram visualization of the multiplication task for participant 1. The yellow colors indicate frequency content with higher power while the blue colors indicate frequency content with very low power.

For each participant under each of the five cognitive task, for each trial, and for each of the six channels (C3, C4, P3, P4, O1, and O2), the task engagement index given by [8] was computed. This resulted in a total of 6 features (task engagement indices) per trial per task. In order to avoid features in greater numeric ranges dominating those in smaller numeric ranges [26] each feature was linearly scaled to the range (0, 1).

Task Engagement Index

$$= \frac{\text{combined beta power}}{\text{combined alpha power} + \text{combined theta power}} \quad (1)$$

C. CLASSIFICATION

First, data for each participant were classified separately. Six separate feedforward ANN each with single hidden layer trained by backpropagation were designed to classify the five mental tasks (baseline, multiplication, letter, geometric figure rotation, and visual counting) for each participant. Second, a single feedforward ANN was designed to classify the five tasks for all participants. The trainbr network training function in MATLAB was used for all the networks. Trainbr generally works best when the network inputs and targets are scaled so that they fall approximately in the

range $(-1, 1)$ [22]. It updates the weight and bias values according to Levenberg-Marquardt optimization. An ANN may suffer from overfitting, a situation where the network memorizes the training samples, but performs poorly on new datasets. Regularization may be used to improve ANN generalization and avoid overfitting. The trainbr network function used in the present study uses Bayesian regularization which takes place within the Levenberg-Marquardt algorithm. It minimizes a combination of squared errors and weights, and then determines the correct combination in order to produce a network that generalizes well [22]. The performance function used for all networks was the mean square errors (MSE). The MSE is the distance between the ANN's estimate of the test data and the actual test data. The data for each participant were analyzed separately. The default dividerand function, which randomly separates targets into three sets (training, validation, and testing), was used for all networks. For the training of each network, the algorithm was allowed to run until the effective number of parameters has converged. Each training stopped with the message "Maximum MU reached."

The best performing single layer feedforward ANN (with least MSE) for each participant was determined by varying the number of hidden nodes from 10 to 50 in steps of 5. A total of ten ANN was analyzed for each participant. Out of the ten ANN analyzed, the one with the least MSE was chosen and its classification accuracy was obtained.

To determine the best performing ANN (with least MSE) for all participants, the number of hidden nodes of single layer feedforward ANN and two layer feedforward was varied from 10 to 50 in steps of 5 for the first layer, and from 5 to 20 in steps of 5. For each n-layer ($n=1$ and 2) feedforward ANN, a total of forty ANN was analyzed. Classification accuracy for the best performing ANN was obtained.

TABLE 1. CLASSIFICATION RESULTS OF 1-LAYER ANN FOR EACH PARTICIPANT

Participant	Number of Hidden Nodes	Epoch	MSE	Classification Accuracy
1	50	89	3.80445×10^{-16}	100
2	45	4750	0.0821047	72
3	40	115	6.9036×10^{-17}	100
4	40	125	1.27205×10^{-12}	100
5	20	311	2.18414×10^{-16}	96
6	30	5238	0.122281	64
Average Classification Accuracy				88.67

III. RESULTS AND DISCUSSION

Table 1 shows the classification results of the single layer ANN for each participant. The classification accuracy of 64% for participant 6's ANN with 30 neurons in its hidden layer was the highest. It also had the highest number of epochs, and highest MSE of 0.122281. The classification accuracy of the ANN for participants 1, 3, and 4 is 100%.

The ANN for participant 5 produced the lowest classification accuracy of 62% with 50 neurons in its hidden node. The average classification accuracy across all participants was 72.3%.

The performance plot and network diagram of ANN for participant 1 are shown in figures 4a and 4b respectively.

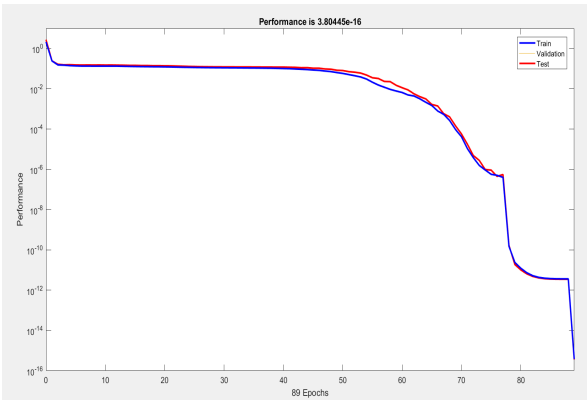


Fig. 4a. Performance plot of ANN for participant 1.

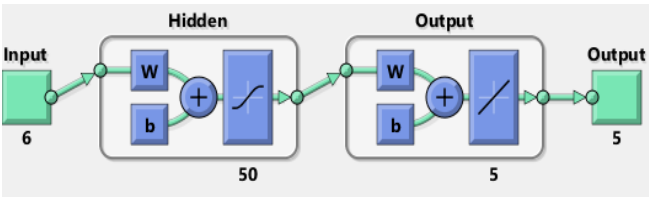


Fig. 4b. Network diagram of ANN for participant 1.

The confusion matrix for the ANN designed for participant 1 and participant 2 are shown in figures 5a and 5b respectively.

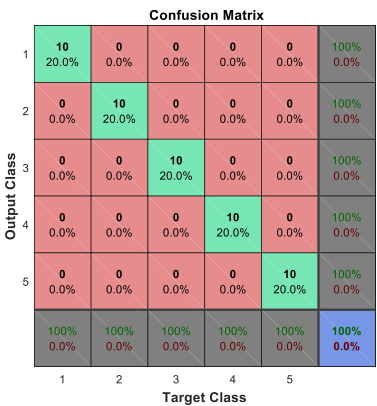


Fig 5a. Confusion Matrix for ANN for participant 1.

Confusion Matrix						
Output Class	1	2	3	4	5	
	10 20.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	0 0.0%	10 20.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	0 0.0%	0 0.0%	10 20.0%	0 0.0%	0 0.0%	100% 0.0%
	0 0.0%	0 0.0%	0 0.0%	10 20.0%	0 0.0%	100% 0.0%
	0 0.0%	0 0.0%	0 0.0%	0 0.0%	10 20.0%	100% 0.0%
Target Class						
	1	2	3	4	5	
	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%

Fig. 5b. Confusion Matrix for ANN for participant 2.

Task engagement is defined as “effortful striving towards task goals” (Matthews et al., 2002). The task engagement index measures how cognitively engaged a person is in a task. In the present study, in a within subject experiment, each participant was presented with five different cognitive tasks. Each task was expected to elicit a different degree of mental engagement. For each participant, task engagement indices were used as inputs to an ANN to allow identification and classification of mental engagement. The results show that differences in cognitive task demand do elicit different degrees of mental engagement, which can be measured through the use of the task engagement index. The average classification accuracy across the six participants was 88.67%.

The classification accuracies for participants 2 and 3 were relatively low. This is partly because an ANN is data driven and is prone to small datasets. Both participants had the least datasets of size 25. In addition, it took a relatively longer time for their networks to converge, 4750 epochs for participant 2, and 5238 for participant 3. The datasets for the remaining ANN were relatively large; 50 for participants 2, 3 and 5, and 75 for participant 4.

IV. CONCLUSION

Differences in cognitive task demand elicit different degrees of mental engagement. These differences can be measured through the use of the task engagement index which is correlated with task demands. The task engagement index may serve as features for identification and classification of mental engagement of cognitive tasks.

References

- [1] Hancock, P. A., Jagacinski, R. J., Parasuraman, R., Wickens, C. D., Wilson, G. F., & Kaber, D. B. (2013). Human-automation interaction research past, present, and future. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 21(2), 9-14.
- [2] Christensen, J. C., Estep, J. R., Wilson, G. F., & Russell, C. A. (2012). The effects of day-to-day variability of physiological data on operator functional state classification. *NeuroImage*, 59(1), 57-63.

- [3] Wilson, G. F., & Russell, C. A. (2007). Performance enhancement in an uninhabited air vehicle task using psychophysiological determined adaptive aiding. *Human factors: the journal of the human factors and ergonomics society*, 49(6), 1005-1018.
- [4] Gevins, A., & Smith, M. E. (2006). Electroencephalography (EEG) in neuroergonomics. *Neuroergonomics: The brain at work*, 15-31.
- [5] Fu, S., & Parasuraman, R. (2006). Event-related potentials (ERPs) in neuroergonomics. *OXFORD SERIES IN HUMAN-TECHNOLOGY INTERACTION*, 32.
- [6] Johnson, A., & Proctor, R. (2013). *Neuroergonomics: A cognitive neuroscience approach to human factors and ergonomics*. Palgrave Macmillan.
- [7] Parasuraman, R., & Rizzo, M. (Eds.). (2006). *Neuroergonomics: The brain at work*. Oxford University Press.
- [8] Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological psychology*, 40(1), 187-195.
- [9] Freeman, F. G., Mikulka, P. J., Prinzel, L. J., & Scerbo, M. W. (1999). Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. *Biological psychology*, 50(1), 61-76.
- [10] Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T.,... & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, 78(Supplement 1), B231-B244.
- [11] Szafir, D., & Mutlu, B. (2012, May). Pay attention!: designing adaptive agents that monitor and improve user engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 11-20). ACM.
- [12] Szafir, D., & Mutlu, B. (2013, April). ARTful: adaptive review technology for flipped learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1001-1010). ACM.
- [13] Huang, J., Yu, C., Wang, Y., Zhao, Y., Liu, S., Mo, C., ... & Shi, Y. (2014, April). FOCUS: enhancing children's engagement in reading by using contextual BCI training sessions. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 1905-1908). ACM.
- [14] McMahan, T., Parberry, I., & Parsons, T. D. (2015, February). Evaluating Electroencephalography Engagement Indices During Video Game Play. In *Proceedings of the Foundations of Digital Games Conference*.
- [15] King, L. M., Nguyen, H. T., & Lal, S. K. L. (2006, August). Early driver fatigue detection from electroencephalography signals using artificial neural networks. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE* (pp. 2187-2190). IEEE.
- [16] Khare, V., Santhosh, J., & Anand, S. (2008, December). Performance comparison using five ANN methods for classification of EEG signals of two mental states. In *2008 Annual IEEE India Conference* (Vol. 1, pp. 7-10). IEEE.
- [17] Chai, R., Tran, Y., Craig, A., Ling, S. H., & Nguyen, H. T. (2014, August). Enhancing accuracy of mental fatigue classification using advanced computational intelligence in an electroencephalography system. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 1318-1341). IEEE.
- [18] Belakhdar, I., Kaaniche, W., Djmel, R., & Ouni, B. (2016, March). Detecting driver drowsiness based on single electroencephalography channel. In *2016 13th International Multi-Conference on Systems, Signals & Devices (SSD)* (pp. 16-21). IEEE.
- [19] Keirn, Z. A., & Aunon, J. I. (1990). A new mode of communication between man and his surroundings. *IEEE transactions on biomedical engineering*, 37(12), 1209-1214.
- [20] cs.colostate.edu[Internet]. Fort Collins, Colorado: Brain-Computer Interfaces Laboratory;[cited 2016 July 1]. Available from: http://www.cs.colostate.edu/eeg/main/data/1989_Keirn_and_Aunon
- [21] Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. *Electroencephalogr Clin Neurophysiol*, 52(3).
- [22] MATLAB 2016a, The MathWorks, Inc., Natick, Massachusetts, United States.

- [23] Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1), 9-21.
- [24] Chaumon, M., Bishop, D. V., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of neuroscience methods*, 250, 47-63.
- [25] Anderson, C. W., Devulapalli, S. V., & Stolz, E. A. (1995). EEG signal classification with different signal representations. In *Neural Networks for Signal Processing [1995] V. Proceedings of the 1995 IEEE Workshop* (pp. 475-483). IEEE.
- [26] Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). *A practical guide to support vector classification*.

STUDY 2-2

NEURAL CORRESPONDENCE TO HUMAN COGNITION FROM ANALYSIS TO INTUITION – IMPLICATIONS OF DISPLAY DESIGN FOR COGNITION

Neural Correspondence to Human Cognition from Analysis to Intuition – Implications of Display Design for Cognition

Psychophysiological measures can be used to determine whether a particular display produces a general difference in brain function. Such information might be valuable in efforts to improve usability in display design. In this preliminary study, we aimed to use the electroencephalography (EEG) task load index (TLI), given by the ratio of mean frontal midline theta energy to mean parietal alpha energy, to provide insight into the mental effort required by participants performing intuition-inducing and analysis-inducing tasks. We employed behavioral measures (reaction time and percent correct), and a subjective measure (NASA-Task Load Index) to validate the objective measure (TLI). The results we obtained were consistent with our hypothesis that mental effort required for analysis-inducing tasks would be different from that required for intuition-inducing tasks. Although our sample size was small, we were able to obtain a significant positive correlation between NASA-Task Load Index and TLI.

INTRODUCTION

Patterson (2017) provides evidence for dual-process cognition through review and analysis of five (heuristics and biases, dual-process theory, fuzzy-trace theory, naturalistic decision making, automatic processes) literature on human decision making. Previous researchers have used dual-process theories to explicate the human cognitive system. The theories supported by much evidence in cognitive science (Evans & Stanovich, 2013), have been the focus of contemporary research (Patterson, 2017; Evans & Stanovich, 2013; Kahneman, 2011; Hammond, 2007; Epstein, 2008), and have had various labels attached to each of one of them. Labels include analysis and intuition (Hammond, 1996), experiential system and rational system (Epstein, 2008), and system 1 and system 2 (Kahneman, 2011). According to Evans (2012), the dual-system theories are basically the same. Evans and Stanovich (2013) refer to these as Type 1 and Type 2. Type 1 processing is autonomous and does not require working memory. On the other hand, Type 2 processing supports hypothetical thinking and requires working memory. Type 1 processing is typically relatively fast, nonconscious, associative, and is independent of cognitive ability. In contrast, type 2 processing is typically slow, conscious, rule-based, and correlated with cognitive ability.

Analysis cognition "...signifies a step-by-step, conscious, logical defensible process..." (Hammond, 1996, pp. 60), that is slow in data processing, high in cognitive control, has a task-specific organizing principle, and has high confidence in the data processing method but low confidence in answer (Hammond, Hamm, Grassia & Pearson, 1987). Intuition cognition, on the other hand, is characterized by low cognitive control, rapid rate of data processing, low conscious awareness, weighted average organization principle, high confidence in answer but low confidence in method (Hammond et al., 1987). Hammond et al. (1987) present task properties that differentially induce intuition and analysis. The properties include number of cues, measurement of cues, distribution of cues, redundancy of cues, decomposition of task, degree of certainty of task, relation

between cues and criterion, weighting of cues in environmental model, availability of organizing principle, display of cues, and time period.

Traditional behavioral and subjective methods employed by system designers to evaluate human-automation interactions may be inadequate (Frey, Mühl, Lotte, & Hachet, 2013). Physiological signals are attractive because they are continuously available and their collection does not interfere with human operators' job performance (Wilson & Russell, 2003). Physiological measures include electroencephalography (EEG), heart rate, eye movements, pupil size, and electrodermal measures. The physiological information, along with behavioral and subjective assessments, can then be used to determine whether the modified system produces same human responses as compared to a baseline system (Kramer & Weber, 2000). Physiological measures can be used to determine whether a particular display or control device produces a general difference in brain function. Consequently, it should be possible to observe changes in a HO's brain signals in order to examine his or her cognitive processes in response to intuition- and analysis-inducing tasks.

Hogervorst, Brouwer and van Erp (2014) combined and compared EEG, peripheral physiology and eye-related measures for the assessment of mental workload. They found that EEG performed best, and that combining different sensors did not significantly improve workload assessment. EEG is well suited for human factors (Gevins & Smith, 2006). EEG is a psychophysiological technique for studying brain activation. EEG signals represent summed postsynaptic potentials of neurons firing a rate of milliseconds. Graphically, an EEG is a graph of the time varying voltage difference between an active electrode attached to the scalp and a reference electrode (Gevins & Smith, 2006). The frequency bands in an EEG signal are delta (less than 4Hz), theta (4–8Hz), alpha (8–12Hz), beta (12–30Hz), and gamma (30–80Hz). These bands are able to provide insight into a person's cognitive states such as workload.

Smith, Gevins, Brown, Karnik and Du (2001) found that frontal midline theta activity increased and alpha band activity decreased with increased task difficulty. The task load index (TLI) as the ratio of theta activity at frontal midline sites to alpha at parietal sites (Gevins & Smith, 2003). The TLI measures mental effort and cognitive load (Kamzanova, Kustubayeva & Matthews, 2014). Previous researchers have employed the TLI to evaluate workload. For example, Ewing, Fairclough and Gilleade (2016) developed an adaptive game system that utilized real-time EEG to maximize player engagement in accordance with the level of game demand. In the first of two studies, they demonstrated the sensitivity of frontal theta and parietal alpha power to changing game demand levels. Hsu, Wang, Chen, and Chen (2015) found the TLI as one of the most effective physiological indices that accurately monitor mental workload. They also found that the TLI highly correlated with subjective measures across different mental workload levels.

Aim of Study

The brain's neural responses to specific cognitive events should afford us the opportunity to study the neural correlates of intuition and analysis cognition. The aim of this study is to use the TLI, given by the ratio of frontal midline theta to parietal alpha, to provide insight into the cognitive load imposed on participants performing intuition-inducing and analysis-inducing tasks.

We hypothesize that the mental effort required for analysis-inducing tasks would be different from that required for intuition-inducing tasks. We employ behavioral measures (reactive time and percent correct), and a subjective measure (NASA-Task Load Index) to validate the objective measure (TLI).

METHOD

Participants

Ten right-handed participants (graduate and undergraduate students) from a south eastern university in the US participated in the study. Participants (five males, and five females) were aged between 20 and 35. Each participant read and signed the informed consent form. All participants had no history of neuropsychiatric disorders, and were not taking psychoactive medications.

We excluded data from four participants from the analyses because of excessive eye or muscle artifacts. We used data from the remaining six participants, five males and one female (aged 20–35 years, mean age of 22.9 and standard deviation of 4.9), for all analyses.

Apparatus

Participants were fitted with g.GAMMAcap (g.tec®, Austria) which uses the International 10-20 system for EEG electrode placement locations on the scalp. Twenty active AgCl electrodes were used: Fp1, Fp2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, Oz, O2 (Fig. 1). The electrodes were connected and amplified with a g.HIamp (g.tec, Austria) amplifier. We programmed and presented the stimuli using Presentation® software suite (Neurobehavioral Systems). We visually inspected and recorded EEG signals with g.Recorder (g.tec, Austria).

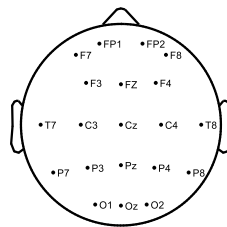


Figure 1. Electrode Placement.

We used a computer version of NASA Task Load Index (Hart & Staveland, 1988) in this study. The NASA TLX is considered one of the most effective measures of perceived mental workload (Wickens, Hollands, Banbury, & Parasuraman, 2015). It provides a measure of overall workload on a scale of 0 to 100 and identifies the relative contributions of six sources of workload. Three of these (mental, physical, and temporal demand) reflect the demand that a task places on a HO. Three others (performance, effort, and frustration) characterize the interactions between observers and the task confronting them. In the present study, we determined workload scores on each of the six subscales using the unweighted scoring procedure recommended by (Nygren, 1991). We performed all EEG preprocessing, and EEG data analyses offline with MATLAB (The MathWorks, 2016), EEGLAB (Delorme & Makeig, 2004), and custom software. We conducted all statistical analyses with SPSS version 24.

Stimuli

We conducted the experiment to measure brain signals while participants performed various cognitive tasks. We designed two task conditions and a baseline. *Baseline.* In this baseline task, we instructed participants to relax and fixate on a blank screen for 60 s.

Intuition-Inducing Task. Stimulus was composed of two objects: a fixation on the left and a flashing image of a human face on the right. We took the faces from the FACE database established by Ebner, and Lindenberger (2010). We instructed participants to press the left mouse button if they thought the face on the right was a happy face, and the right mouse button if they thought the face on the right was the face of someone who was afraid. They were instructed to make their judgment quickly and most importantly, to follow their first feeling. We presented each stimulus for 60s followed by a variable (2~4 s) inter-stimulus interval (ISI) during which we replaced the stimulus with a fixation cross.

Analysis-Inducing Task. Stimulus was composed of two objects: a multiplication of a double digit number by a single digit number on the right and a different multiplication of a double digit number by a single digit number on the left. We instructed participants to press the left mouse button if they thought the multiplication on the left was larger, and the right mouse button if they thought the multiplication on the right was larger. We informed them that they had enough time to make their decisions. We presented each stimulus for 60 s followed by a variable (2~4 s) inter-stimulus interval (ISI) during which we replaced the stimulus with a fixation cross.

Experimental Design

We employed a within-subjects design in the present study. The independent variable was task type (analysis-inducing and intuition-inducing). The dependent variables were EEG recorded as participants perform cognitive tasks, subjective NASA TLX rating administered after each task, mean reaction time, and mean percent correct. From the recorded EEG, we obtained the TLIs for each task type.

Procedure

We asked participants to first sign an informed consent form and then fill out a demographic questionnaire. Next, we fitted them with the g.GAMMAcap. Before recording the EEG, we checked electrode impedance and calibrated it to be less than 5k Ω . We used the right ear lobe as reference. The experiment consisted of three conditions: (1) baseline (eye-opened resting condition), (2) analysis-inducing task, and (3) intuition-inducing task. There were 2 sessions: each session consisted of 1 block of 30 trials for the analysis-inducing condition, 1 block of 30 trials for intuition-inducing condition, and 1 baseline task.

We presented the visual stimuli on a 24" LED monitor (with a resolution 1920 x 1080) that was positioned at the same height as the participant's eyes. To avoid order effects, we presented the block of trials and the trials in a block in pseudorandom order.

We used short blocks of trials in order to encourage participants to remain engaged in the task throughout the experiment. Each intuition-inducing trial had maximum duration of 60s. The trial ended when participants clicked the mouse button before or at the stipulated duration of 60s. A cross fixation of variable duration (2~4s) was presented between trials (see Figure 2).

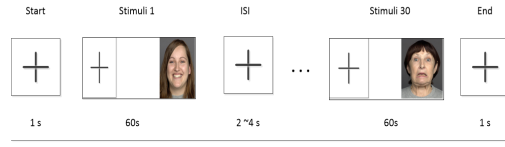


Figure 2. Block of trials for intuition-inducing trials.

Each analysis-inducing trial had maximum duration of 60s. The trial ended when participants clicked the mouse button before or at the stipulated duration of 60s. A cross fixation of variable duration (2~4s) was presented between trials (see Figure 4).

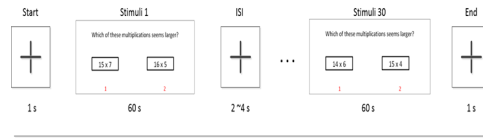


Figure 3. Block of trials for analysis-inducing trials.

There were 2-minute breaks between blocks, and 5 minutes break between sessions. Response time (RT) and number of correct responses were recorded for the analysis-inducing and intuition-inducing conditions. We administered the NASA TLX questionnaire after each block of trials. Finally, we debriefed and thanked participants.

EEG Recording and Preprocessing

We recorded raw EEG signals at a sampling rate of 256 Hz using a Butterworth filter (0.01Hz high pass–100Hz low pass). We applied a notch filter with 60 Hz cutoff frequency to remove line noise. To remove linear trends, we high-pass filtered the data with basic Finite Impulse Response (FIR) filter that had a 1 Hz as the lower edge frequency. We re-referenced the recorded EEG data to average. In the present study, we time locked the EEG epochs to stimulus presentation. We extracted the data epochs time locked to events of interest. Epochs started at 1s before stimulus onset and ended at 2s (see Figure 4).

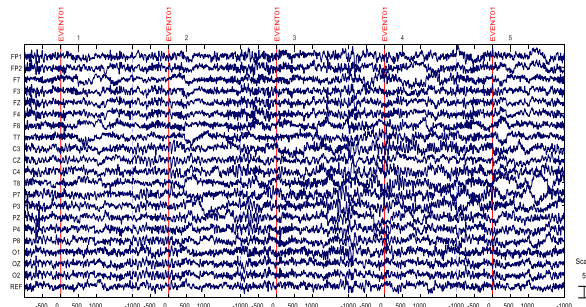


Figure 4. Epoched EEG Signal showing stimulus onset triggers.

We rejected data with amplitudes outside of the range of $-50 \mu\text{V}$ to $+50 \mu\text{V}$. The rejection rate was around 5–10%. We used the SASICA software (Chaumon, Bishop, & Busch, 2015) to reject artifact independent components (ICs) before EEG data analysis.

Wavelet Analysis

We used the continuous wavelet transform (CWT) to extract features from the recorded EEG signals. The CWT is expressed as

$$C(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (1)$$

where $a, b \in \mathbb{R}$, $a \neq 0$, and $\mathbb{R}$ is the set of real numbers. a is the scale parameter, b is the location parameter, $\psi(t)$ is the wavelet function also known as the ‘mother wavelet’.

The superscript ‘*’ denotes the complex conjugate of the function, and $\frac{1}{\sqrt{a}}$ is used to normalize the energy such that it stays at the same level for different values of a and b (Ghorbanian, Devilbiss, Simon, Bernstein, Hess, & Ashrafiuon, 2012).

Wavelet analysis overcomes the limitations of the fast Fourier Transform (FFT) by breaking up of the EEG signal into shifted and scaled versions of the original (or mother) wavelet. The choice of the mother wavelet depends on what kind of features need to be extracted from the signal, and is extremely important in CWT. By this choice, the time and frequency resolution of the result can be influenced. Wavelets that are well localized in the frequency domain, like the Morlet wavelet, are used for detection of salient oscillations (Senkowski & Herrmann, 2002). In the present study, we used the Morlet wavelet. We calculated the coefficients of CWT and computed the geometric mean energy of the wavelet coefficients of each scale.

The mean value of geometric means at each scale gives their corresponding absolute energy, E_{band} . For an EEG signal, the total energy E_t across all five bands is given by

$$E_t = E_{delta} + E_{theta} + E_{alpha} + E_{beta} + E_{gamma} \quad (2)$$

Relative wavelet energy ρ_{band} is defined as (Rosso et al., 2001)

$$\rho_{band} = \frac{E_{band}}{E_t} \quad (3)$$

For each participant, we computed the relative energy features of all the channels’ data. Next, for each participant under each cognitive task, for each trial, we computed two relative wavelet energy features ($\rho_{theta}, \rho_{alpha}$). Then, we averaged frontal theta relative energy features per task. Also, we averaged parietal alpha relative energy features per task. Finally, for each participant and for each task we computed the TLI given by

$$TLI = \frac{Frontal\ Theta}{Parietal\ Alpha} \quad (4)$$

RESULTS

Subjective Workload Assessment

Figure 5 shows the effects of task type on the perceived workload. A Wilcoxon signed-rank test indicated a higher perceived workload for analysis-inducing tasks ($M =$

60.42, SD = 4.16) than (M = 60.42, SD = 4.16) and intuition-inducing tasks, $Z = -2.201$, $p = 0.028$.

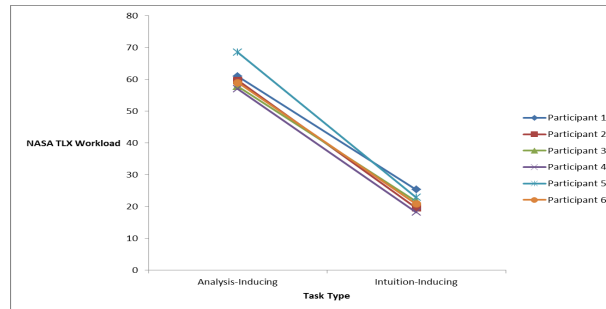


Figure 5. Effect of Task Type on NASA-TLX.

Behavioral Measures: Percent Correct and RT

The effect of task type on percent correct is shown in Figure 6. A Wilcoxon signed-rank test indicated a lower percent correct for analysis-inducing tasks (M = 91.61, SD = 3.07) than (M = 98.60, SD = 1.04) and intuition-inducing tasks, $Z = -2.101$, $p = 0.022$.

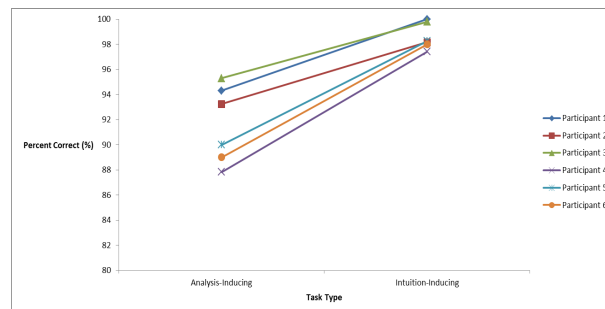


Figure 6. Effect of Task Type on Percent Correct.

The effect of task type on mean reaction time is depicted in Figure 7. A Wilcoxon signed-rank test indicated a longer reaction time for analysis-inducing tasks (M = 22682.19, SD = 7594.32) than (M = 6155.59, SD = 437.89) and intuition-inducing tasks, $Z = -2.507$, $p = 0.015$.

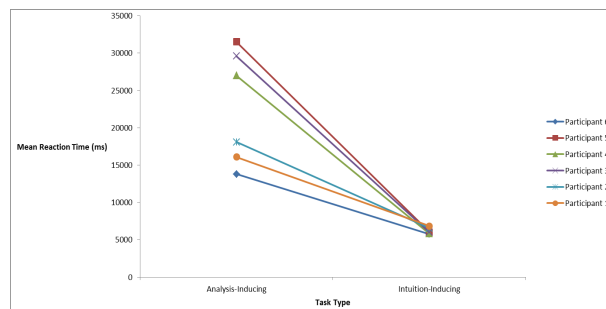


Figure 7. Effects of Task Type on Mean Reaction Time.

EEG Task Load Index

In Figure 8 we show the effect of task type on TLI. A Wilcoxon signed-rank test indicated a higher TLI for analysis-inducing tasks ($M = 5.22$, $SD = 1.73$) than ($M = 3.81$, $SD = 1.61$) and intuition-inducing tasks, $Z = -2.209$, $p = 0.025$.

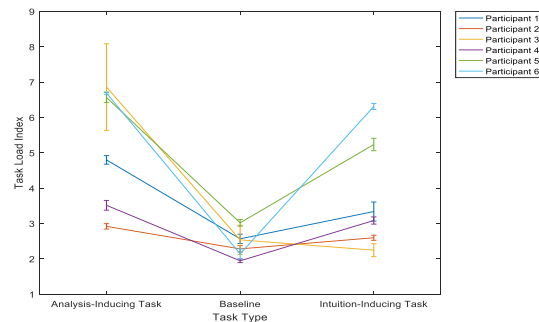


Figure 8. Effect of Task Type on Task Load Index. Error bars are standard errors.

Correlational Analysis

We conducted a Pearson correlation analysis to identify the relationship between NASA TLX and EEG TLI. There was a positive correlation between the NASA-TLX and TLI, $r = 0.194$, $p = 0.04$.

DISCUSSION

The purpose of this study was to determine whether the TLI, given by the ratio of the mean frontal midline theta energy to the mean parietal alpha energy, derived from each participant would succeed in distinguishing between intuition-inducing and analysis-inducing tasks. We hypothesized that mental effort required for analysis-inducing tasks would be different from that required for intuition-inducing tasks. The results we obtained from the statistical analysis were consistent with our hypothesis.

Intuition and analysis are both quantitatively and qualitatively different cognitive processes. Intuition is characterized by low cognitive effort while analysis is characterized by high cognitive effort. As in prior studies (Nuamah & Seong, 2017; Fairclough & Gilleade, 2016; Hsu et al., 2015), mean TLI across participants appeared to be influenced by changes in cognitive load. Results of the subjective workload, assessed with the NASA-TLX, validated subjective TLI measures. Figure 8 shows that both intuition-inducing, and analysis-inducing tasks generated TLI greater than baseline. Although our sample size was small, we were able to obtain a significant positive correlation ($r = 0.194$) between NASA-TLX and TLI.

Acknowledgments

This work was partially supported by DoD Army Research Office (W911NF-13-1-018), and by Air Force Research Laboratory (FA8750-15-2-0116) through the second author.

References

Chaumon, M., Bishop, D. V., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of neuroscience methods*, 250, 47-63.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1), 9-21.

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior research methods*, 42(1), 351-362.

Epstein, S. (2008). Intuition from the perspective of cognitive-experiential self-theory. *Intuition in judgment and decision making*, 23, 37.

Evans, J. S. B. (2012). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning*, 18(1), 5-31.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3), 223-241.

Ewing, K. C., Fairclough, S. H., & Gilleade, K. (2016). Evaluation of an Adaptive Game that Uses EEG Measures Validated during the Design Process as Inputs to a Biocybernetic Loop. *Frontiers in human neuroscience*, 10.

Frey, J., Mühl, C., Lotte, F., & Hachet, M. (2013). Review of the use of electroencephalography as an evaluation method for human-computer interaction. *arXiv preprint arXiv:1311.2222*.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139-183.

Hsu, B. W., Wang, M. J. J., Chen, C. Y., & Chen, F. (2015). Effective indices for monitoring mental workload while performing multiple tasks. *Perceptual and motor skills*, 121(1), 94-117.

Gevins, A., & Smith, M. E. (2006). Electroencephalography (EEG) in neuroergonomics. *Neuroergonomics: The brain at work*, 15-31.

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, 4(1-2), 113-131.

Ghorbanian, P., Devillbiss, D. M., Simon, A. J., Bernstein, A., Hess, T., & Ashrafiuon, H. (2012, October). Continuous wavelet transform EEG features of Alzheimer's disease. In *ASME 2012 5th Annual Dynamic Systems and Control Conference joint with the JSME 2012 11th Motion and Vibration Conference* (pp. 567-572). American Society of Mechanical Engineers.

Hammond, K. R. (2007). *Beyond rationality: The search for wisdom in a troubled time*. Oxford University Press.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE transactions on Systems, Man, and Cybernetics*, 17(5), 753-77.

Hogervorst, M. A., Brouwer, A. M., & van Erp, J. B. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kamzanova, A. T., Kustubayeva, A. M., & Matthews, G. (2014). Use of EEG workload indices for diagnostic monitoring of vigilance decrement. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 56(6), 1136-1149.

Kramer, A. F., & Weber, T. (2000). Applications of psychophysiology to human factors. *Handbook of psychophysiology*, 2, 794-814.

MATLAB 2016a, The MathWorks, Inc., Natick, Massachusetts, United States.

Nuamah, J. K., & Seong, Y. (2017). Support vector machine (SVM) classification of cognitive tasks based on electroencephalography (EEG) engagement index. *Brain-Computer Interfaces*. Advance online publication. doi: 10.1080/2326263X.2017.1338012.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 33(1), 17-33.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, 30(3), 86-297.

Patterson, R. E. (2017). Intuitive Cognition and Models of Human–Automation Interaction. *Human Factors*, 59(1), 101-115.

Rosso, O. A., Blanco, S., Yordanova, J., Kolev, V., Figliola, A., Schürmann, M., & Başar, E. (2001). Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of neuroscience methods*, 105(1), 65-75.

- Senkowski, D., & Herrmann, C. S. (2002). Effects of task difficulty on evoked gamma activity and ERPs in a visual discrimination task. *Clinical Neurophysiology*, 113(11), 1742-1753.
- Smith, M. E., Gevins, A., Brown, H., Karnik, A., & Du, R. (2001). Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction. *Human Factors*, 43(3), 366-380.
- Wilson, G. F., & Russell, C. A. (2003). Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(3), 381-389.

STUDY 3

AN INVESTIGATION OF THE COGNITIVE PROCESS OF HUMAN TRUST IN AUTOMATION WITH NEURAL ACTIVITIES

AN INVESTIGATION OF THE COGNITIVE PROCESS OF HUMAN TRUST IN AUTOMATION WITH NEURAL ACTIVITIES

Abstract

In the contemporary society, autonomous system have become complex because of the strong need for sophisticated tasks, and the need to design complex control systems using advanced technology. However, complex autonomous systems can increase unstable and uncertain situations in workplaces that eventually will lead to increased uncertainty, which will increase cognitive complexity (Lee & See, 2004). In complex autonomous systems such as motor vehicle operation, aviation, maritime operation and process control, the role of the human operator is crucial in controlling and monitoring to avoid failure and risk, and to prevent unpredictable situations that cannot be programmed in the system. Trust can affect the operator's degree of acceptance and reliance on the automated system, so measuring human operators' level of trust is vital in predicting their strategies during interaction with the system. Previous studies that measured human trust highly depended on the use of questionnaires to measure subjective feelings of trust based on a theoretical approach (Larzelere, & Huston, 1980; Rempel, et al., 1985; Singh, Molloy, & Parasuraman, 1993; Lee & Moray, 1994; Muir & Moray, 1996; Bisantz & Seong, 2001). However, many questionnaires used in these studies are based on a theoretical approach that does not have an objective measure and can vary by respective researchers' theoretical orientation. To supplement a theoretical approach, there are questionnaires which used empirically developed trust questionnaire (Jian et al., 2000). There are other researches, which are still using questionnaires, but they are used with simulated models. Due to the limitations of the use of questionnaire surveys based on theoretical or empirical approaches used to measure trust, this research proposes to use a neurological measure, which is a direct and objective method to detect and measure human trust.

Neuroimaging techniques can assist to examine various cognitive processes including language, reasoning, problem solving, decision making and other complicated mental functions (Burge & Kahn, 2009). Previous studies investigated the correspondence of relationship between social constructs (i.e., trust) and neural or physiological evidence and focused on brain regions using fMRI and eMRI, which engage with stimuli related to trust and distrust. However, the previous studies did not fully investigate brainwaves in the situations of trust and mistrust. This research proposes to measure and analyze human brainwaves using situations involving trust and mistrust to understand the cognitive process using an electroencephalogram (EEG). EEG can detect the response to the external stimuli through recording the brain's electrical activity in a real time, so it is useful for scientific experiments with a wide range of cognitive tasks.

The overall purpose of the proposed research within the domain of human trust in the use of automation is to understand how the cognitive process of trust and mistrust affect decision making, which is the human operator's use of automation. Therefore, the objectives of the proposed research are to investigate the human cognitive process using EEG by recording, identifying and analyzing specific brainwaves and brain regions in situations involving trust and mistrust, and to discover how the cognitive process of trust and mistrust effects the human operator's use of automation. More specifically, the proposed study seeks to conduct experiments of two phases: Phase I and Phase II. The

Phase I experiment will use a word elicitation study to identify specific brainwaves and brain regions resulting from situations of trust and mistrust. The Phase II experiment will use a simulation to evaluate strategies of individual human operator's strategy on the use of automation in the situations of trust and mistrust. Eventually the proposed research will investigate the cognitive processes of trust and mistrust by analyzing corresponding data among identified brainwaves, stimulated brain regions, and the strategies of individual human operator's use of automation. Supported by the expected result, trust in automation will assist in effective decision-making by increasing concentration and performance, but mistrust in automation will interrupt effective decision-making by increasing stress and anxiety. Therefore, the findings of this research will contribute to defining how trust in automation affect the human operator's decision-making and overall performance.

CHAPTER 1. INTRODUCTION

An Autonomous system is the use of various control systems for operating equipment with minimal or reduced human intervention. It has been actively used in many applications such as machinery, factories, ships, aircraft and vehicles to consistently improve quality, accuracy, productivity and efficiency. Automation can also reduce direct human labor costs and expenses by replacing humans in tasks that involve physically hard and monotonous work and improve safety of tasks in dangerous and hazardous environments. Automated systems have evolved very rapidly into complex sophisticated tasks and advanced technology. Trust can be a 'decision aid' in dynamic and uncertain situations when complex automated systems are impossible to comprehend and when flexible behaviors are necessary to combat unexpected situations that arise, which are unavoidable using procedures (Lee & See, 2004). Trust also plays a crucial role in contributing to the cognitive complexity and increased uncertainty in sophisticated automation systems.

To understand trust between human and automation, it is necessary to study how to define trust. Trust has been defined related to expectation. Deutsch (1973) defined trust as "confidence that one will find what is desires from another" and Rempel, Holmes, and Zanna (1985) defined it as "a generalized expectation related to the subjective probability" to future events. A sociologist Barber (1983) defined trust with three specific expectations such as persistence, technical competence, and fiduciary responsibilities. By integrating the study of Barber (1983) and Rempel et al. (1985). Muir (1987) extended her study to develop trust in human-machine relationships and claimed that trust is developed by "the human's ability to estimate the predictability of the machine's behaviours." Lee and Moray (1992) developed dimensions of trust, which includes a taxonomy for trust from the research of Barber (1983), Rempel, et al. (1985), Muir (1989) and Zuboff (1988). They presented the foundation, performance, process and purpose as the basic dimensions to understand trust.

Performance of trust includes current and past operations categorized as reliability, predictability, ability, competence and consistency.

Several researchers investigated factors that influence the use of automation, and found that trust played an important factor that affects misuse and disuse of automation (Lee & Moray, 1992; Muir, 1987; Bisantz & Seong, 2001; Lee & See, 2004). Trust is a variable to decide the use of automation (Masalonis & Parasuraman, 1999). Measuring the human operator's level of trust in automation becomes crucial in predicting their strategies while using automation.

However, trust is a multidimensional construct influenced by various factors such as motives, intentions, and actions" (Bhattacharya, Devinney, & Pillutla, 1998; Mayer et al., 1995), so it is challenging to measure trust.

As a method to measure trust, questionnaires are mostly used to investigate trust in interpersonal relationships (Larzelere, & Huston, 1980; Rempel, et al., 1985) with an empirical approach (Jian, Bisantz & Drury., 2000; Bisantz, Llinas, Seong, Finger, & Jian, 2000) and with a semi-automatic simulation of human and automation (Lee, & Moray, 1994). Further, the questionnaires used in these studies are based on a theoretical or empirical approach, which can vary by the researcher's respective theoretical orientation or specific model of simulations.

1.1 Background

The origin of trust comes from *trost*, which means comfort in German. If I trust someone, it means that I am comforted by the belief that he or she will not do any action contrary to my expectation. Many researchers proposed to define trust between humans related to expectation. Deutsch (1973) defined trust as "confidence that one will find what is desired from another rather than what is feared. Scanzoni (1979) claims trust as "an actor's willingness to arrange and repose his or her activities on others because of the confidence that others will provide expected gratification." Rotter (1980) defined the concept of interpersonal trust as "a generalized expectancy held by an individual that the world promise, oral or written statement of another individual or group can be relied on." Rempel, Holmes, and Zanna (1985) explained trust as "a generalized expectation related to the subjective probability of an individual assign to the occurrence of some set of future events" and later Rempel and Holmes (1986) defined trust as "the degree of confidence one feels when thinking about a relationship." Barber (1983) understood the multidimensional concept of trust and defined trust with three expectations of "the persistence of the natural and the moral social orders, technically competent role performance, and fiduciary obligations and responsibilities." By combining Barber (1983) and Rempel et al (1986), Muir (1987) defined trust in human-machine relationship as "trust is the expectation, held by a member of a system, of persistence of the natural and moral social orders, and of technically competence performance, and of fiduciary responsibility, from a member of the system, and is related to, but not necessarily isomorphic with, objective measures of these qualities." Therefore, it is necessary to understand trust as a multidimensional concept. Trust is highly related to the idea of social influence because it would be much easier to influence or persuade someone who is trusting. Trust can also have a positive influence on a person's behaviors, perceptions and performances. Trust can predict the acceptance of behaviors by other individuals, social groups such as friends, communities, organizations, companies, nations, and

objects such as machines and automated systems.

Trust between humans and machines is similar. Trust is developed over time but very hard to regain once gone. Humans (trustor) can develop trust in a machine (trustee) by evaluating the performances of the machine over time. With trust, then human (trustor) is willing to rely on the use of the machine (trustee). If the machine's outcome is unexpected and unstable involving the risk of failure or harm to humans (trustor), humans (trustor) will lose their trust and not rely on the machine in the future. Building trust is also influenced by previous experiences, the current state is to believe and the future expectation is to have faith. Therefore, the key here is maintaining a level of trust in the interaction between humans and machines.

1.2 Problem Statement

In the interaction between humans and automation, it is important to measure human operator's level of trust in automation to predict the strategy of the operator in the use automation. Questionnaires are widely used to investigate and measure trust . Larzelere and Huston (1980) established a dyadic trust scale using questionnaire surveys between partners. Rempel, et al (1985) developed a trust scale within close interpersonal relationships with subscales of predictability from past experiences, dependability about partner's personality, and faith from the feeling of confidence and expectation for the future. Previous studies focused on specific individuals such as romantic partners; therefore, these trust scales have limitations that apply to general people or objects such as automation. The method of questionnaires can measure trust as a multidimensional concept with a set of questions to examine various aspects of trust, but it based on a theoretical approach, which do not have an objective standard and can vary by the researchers' respective theoretical orientation.

To supplement with the theoretical questionnaires and the study of trust between humans, Lee and Moray (1994) used an experimental system such as a semi-automatic pasteurization plant to investigate the change in an operator's control strategies and trust during interaction with the automation. The operator had the two goals of performance and safety, and made decisions by using manual control, automatic control or a combination of the two. They evaluated trust based on data collected from subjective but specific questions about the operator's trust in automation and the self-confidence in specific parts of automation. The study used a well-designed experiment of 60 trials with a training trial and the occurrence of fault that examined various aspects of human-machine trust that could influence the operator's strategies. This study still uses the method of questionnaires, but it uses empirical questionnaires based on the actual interaction between humans and machine after each trial. This particular study used an experimental system with specific tasks and complexity to measure trust, so it can limit to apply the measure trust to other automated systems, which have different tasks and level of complexity, or individuals, who have different previous skills and knowledges.

1.3 Research Gaps

Measuring the human operator's level of trust in automation becomes crucial in predicting their strategies while using automation. How can we measure human trust in the use of automation effectively? Many researchers used questionnaires that measure subjective feelings of trust (Larzelere, & huston, 1980; Rempel, et al., 1985; Singh,

Molloy, & Parasuraman, 1993; Lee & Moray, 1994; Muir & Moray, 1996, Bisantz & Seong, 2001).

However, many questionnaires used in these studies are based on a theoretical approach that does not have an objective measure and can vary by respective researchers' theoretical orientation. To supplement a theoretical approach, there are questionnaires which used empirically developed trust questionnaire (Jian et al., 2000). There are other researches, which are still using questionnaires, but they are used with simulated models.

Due to the limitations of the use of questionnaire surveys based on theoretical or empirical approaches used to measure trust, this research proposes to use a neurological measure, which is a direct and objective method to detect and measure human trust.

Thus far, some studies have investigated the correspondence of relationship between social constructs (i.e., trust) and neural or physiological evidence. Winston, Stranger, O'Doherty, and Dolan (2002) investigated the correlation between the degree of trust and human facial appearance using event-related functional magnetic resonance imaging (efMRI). They found that trustworthiness was correlated in a frontopolar cortex and untrustworthiness was correlated with right amygdala and right insula. Krueger, McCabe, Moll, Kriegeskorte, Nikolaus, Zahn, Strenziok, Heinecke, Armin, and Grafman, (2007) examined interpersonal trust when two strangers interacted online with one another in a trust game and detected different brain regions based on conditional and unconditional trust in a partnership using hyper-fMRI. They discovered that conditional trust activated the ventral tegmental area while unconditional trust activated the septal area. Dimoka and Angelika (2011) had an experiment to find a correlation between brain regions and trust using fMRI by measuring response of two fictitious online sellers who have high and low reputation. They revealed found a correlation between brain regions and trust that trust is related to the caudate nucleus, putamen, anterior paracingulate cortex, and orbitofrontal cortex while distrust is related to the amygdala and insular cortex. These previous studies focused on brain regions using fMRI and efMRI, which engage with stimuli related to trust and distrust. However, the previous studies did not investigate brainwaves in the situations of trust and mistrust.

This research proposes to measure and analyze human brainwaves using situations involving trust and mistrust to understand the cognitive process using an electroencephalogram (EEG). Electroencephalography (EEG) is an electrophysiological monitoring method that records the electrical activity of the brain by placing electrodes on the scalp. EEG has several benefits for measuring brain activity. EEG has a wide range of intra-individual and inter-individual variability which can be divided in frequency bands (delta, theta, alpha, beta, gamma) (D'Avanzo, Tarantinob, Bisiacchib & Sparacinoa, 2001). EEG detects changes in the brain's electrical activity in response to external stimuli in various regions within a millisecond timeframe in real time. EEG can be sensitively changed to the external stimuli, so it is useful for experiments with a wide range of cognitive tasks, which requires to observe responses by external stimuli. In addition, EEG provides high resolution without exposure to radiation or magnetic fields and it is not expensive compared to other equipment types, portable, and is not claustrophobic.

1.4 Research Objectives

The overall purpose of the proposed research within the domain of human trust in the use of automation is to understand how the cognitive process of trust and mistrust effects the human operator's use of automation. Therefore, the objectives of this dissertation are to investigate the human cognitive process using an electroencephalogram (EEG) by recording, identifying and analyzing specific brainwaves and brain regions in situations involving trust and mistrust, and to discover how the cognitive process of trust and mistrust effects the human operator's use of automation. More specifically, the detailed objectives are as follows.

- This research will identify specific brainwaves in the situation of trust and mistrust.
- This research will identify specific brain regions stimulated in the situations of trust and mistrust.
- This research will evaluate human operator's strategies used during interaction with automation in the situations of trust and mistrust.

This research will investigate the cognitive process of trust and mistrust by analyzing the correspondence between the human operator's trust and neurological activities.

1.5 Research Contributions

This research proposes that trust in automation will assist in the effective decision-making by increasing concentration and performance; nevertheless, mistrust in automation will interrupt the operator's effective decision-making by increasing stress and anxiety. Findings from this research will contribute to defining how trust in automation can affect the human operator's decision-making and overall performance.

1. The proposed research aims to identify specific brainwaves associated with situations of trust and mistrust using an EEG. This research is unique to previous literature because the proposed method provides a direct and objective way to detect and measure human trust and mistrust any assigned task in real time. Current studies focused on the multidimensional concept of trust; however, the proposed research will examine the cognitive process of trust and mistrust, which are unconscious and hidden that needs an in-depth investigation.
2. The proposed research can also be beneficial in designing automated systems, which require gaining the trust of human operators to achieve maximum performance and safety. Automated systems have become complex so they can execute high level tasks; thus, these systems have become difficult to fully understand, properly use and critical for human operators to monitor. By detecting human operator's trust in automation, designers can improve automated systems by developing a user-friendly interface and effective training, which can decrease workload and increase trust.
3. Further, the proposed research can be a means of monitoring the psychological state of the human operators. Human operators can monitor the automated systems to avoid failure and malfunction. The human operators can report their ability to operate the automatic system properly; however, there

are no system or program to monitor the human operators. This EEG method can monitor the human operator and detect the human operator's trust in automation. If the human operator mistrust automation, training or improving the automated system are ways to regain trust. If the human operator is unstable and unfocused due to extreme stress and anxiety, the recommendation is to operate in full automation rather than manual operation, especially in emergency situations.

CHAPTER 2. LITERATURE REVIEW

2.1 Overview

This chapter presents relevant literature on human trust in automation and the human brain functions relevant to the cognitive process. These studies will serve as the underpinnings for the proposed research. The first part will investigate human-machine trust on the appropriate use of automation with current studies. The second part will investigate the cognitive process by examining the human brain regions, EEG analysis, brain waves, and other brain imaging methodologies. Furthermore, a clear connection to measure human trust using EEG will be articulated within the literature review.

2.2 Humans and Automation

Human operator's trust in the complex automation plays an important role in decision making and solving problems during unexpected and uncertain situations. Previous knowledge, training and self-confidence can also affect an operator's strategies when using automation. If the operator mistrust automation, he or she will tend to use manual operation in possible emergent situations. Using manual operation can also increase the operator's workload needed to process all information without the assistance of automation, which will inevitably decrease quality, productivity and efficiency. If the operator over trust automation, he or she cannot correctly monitor and properly control the system. Nor can the operator find the cause of the problem(s). Subsequently, these type situations become difficult to understand and to solve the problem that can cause a serious accident or failure. Hence, trust is an important factor in using automation properly. Research also supports these notions. Uggirala, Gramopadhye, Melloy, and Toler (2004) states the degree of trust can influence system performance because trust can affect the operator's degree of acceptance and reliance on automation (Sheridan, 1988) and the operator's strategies regarding the use of automation (Lee & Moray, 1994).

Proper use of automation can reduce workload and error, and improve performance and safety. However, automation is often misused, disuse or abused because it is influenced by complicated factors such as workload, cognitive overload, trust, confidence and risk (Lee, 2008). Muir (1987) stressed the importance of calibrating human trust in the machine by using automation to achieve good performance. Well calibrated trust can lead to appropriate use of automation.

2.2.1 Misuse of automation.

Misuse is the "overreliance on automation" when it should not be trusted, which can affect automation bias or monitor failure (Parasuraman & Riley, 1997). When operators over-trust automation, they are less attentive to judge information and situations (Skitka, Mosier, & Burdick, 1999, 2000). Monitor failures can be a significant problem in the use of automation because when operators do not pay attention to or ignore automation malfunctions, this can lead to serious incidents. One well-known example is when Eastern Airlines Flight 402 crashed into the Florida Everglades (NTSB, 1973). In this accident, the flight crew failed to recognize that the autopilot was disengaged and did not monitor the altitude correctly because they were fixated on the landing gear indicator

light (NTSB, 1973). A similar example is found in the automated navigation system in ships (Fahey, 2007). The navigation team failed to monitor the automated navigation because they trusted only information provided by the automated system even when the information they observed directly was contradictory to the information. Even though the general system of automation is trustworthy, the automation can have an error or failure. The important role of human operators is to properly monitor automation to avoid the issue of misuse of automation.

2.2.2 Disuse of automation.

Disuse occurs when operators do not utilize automation enough; this can often occur in alarms and safety systems. It can be similar as the phenomenon *The Boy Who Cried Wolf Effect*, which people become tired of hearing warnings and desensitized about the sign of the danger. The rate of false alarms affects reliance on automation. When there is no alarm of automation, participants over relied on the automation. But, when false alarms of automation are dominant, participants under relied on the automation (Sanchez, Rogers, Fisk & Rovira, 2014). If an alarm system is too sensitive to a set condition, this creates false alarm problems; therefore, operators develop mistrust in the alarm systems and eventually do not use them (Parasuraman & Riley, 1997). Alarms are designed to prepare for emergencies, however, repeated minor alarms for non-emergency situations cause alarm fatigue that will cause operators to start ignoring most alarms because they will assume that each time the alarm will probably be not urgent or emergent. When operators ignore the alarm systems in emergency situations; hence, serious accidents can occur. For example, consider the train collision in eastern Baltimore Maryland in 1987. Amtrak Train 94 ran a light, and crashed into a set of Conrail locomotives. This accident would have been avoided if the Conrail locomotive crew had not ignored the alert system (NTSB, 1988). On the other hand, if an alarm system never gives minor alarms in its respective setting, the alarm system probably will not provide any warning or preparation before unexpected incidents. In other words, if an alarm system is too strict to a set condition, operators cannot properly respond to dangerous situations, because it can be the first time to hear the alarm (Farber & Paley, 1993). It is highly recommended to have a proper alarm rate, which can help human operators prepare for emergency situations.

2.2.3 Abuse of automation.

Abuse of automation can be defined as an inappropriate application of automation by designers and managers without considering the consequences for human operators' performance and authority (Parashuraman & Riley, 1997). Abuse of automation often occurs when designers or managers depends on automation over human operators. Consequently, abuse of automation can result in misuse or disuse of automation by human operators. Automation abuse occurs when automation designers give a high degree of authority and autonomy to automation over operators (Starter & Woods, 1994). Automation is generally used to decrease human errors and increase productivity and performance. Parasuraman and Riley (1997) mentioned that by design, automation reduces human errors; however, simply removing human operators from workplaces will not solve problems. To eliminate completely human operators in the design of automation is problematic, it shows that lack of trust in human operators and overtrust in automation. Abuse of automation occurs when human operators have limited

access to automation by managers or policies even emergency situations (Parasuraman & Riley, 1997). Who can monitor when automation has errors that can cause serious accidents? Specifically, in emergency situations unexpected and dynamic, require operators to have flexible responses rather than follow responses of automation that are programmed for limited situations. If human operators can control automation in emergency situations, they can understand the problems, monitor the automation system, recommend solutions and possibly solve the problems. Abuse of automation occurs when automation is designed with the technology-centered approach, which can be difficult to human operators to monitor. If automation is too complicated to understand, it will increase human operators' workload even when it is already high. In this situation, operators can abuse automation because they cannot monitor or respond correctly (Parasuraman & Riley, 1997). Therefore, for avoiding abuse of automation it is necessary to understand the interaction between human operators and automation rather than overtrust or ignoring one of them.

2.3 Human Trust with Appropriation

Several researchers investigated factors that influence the use of automation, and found that trust is an important factor that affects misuse and disuse (Lee & Moray, 1992; Muir, 1987; Lee & See, 2004; Seong & Bisantz, 2008). To investigate trust on automation, it is necessary to understand trust with various aspects and to investigate how trust affects human operators' strategies. Rempel, Holmes, and Zanna (1985) defined trust with predictability, dependability and faith, and Barber (1983) defined trust with three specific expectations such as persistence, technical competence, and fiduciary responsibilities. Muir (1987) developed trust in human-machine relationships and claimed that trust is developed by "the human's ability to estimate the predictability of the machine's behaviours."

Lee and Moray (1992) developed dimensions of trust, which is a taxonomy for trust by Barber (1983), Rempel, et al (1985), Muir (1989) and Zuboff (1988). They presented the foundation, performance, process and purpose as the basic dimensions to understand trust. Foundation of trust refers to the basic principles of natural and social order such as persistence of natural laws (Barber, 1983). Performance of trust includes current and past operations categorized as reliability, predictability, ability, competence and consistency. Process of trust is about understanding the appropriate degree of automation that control performance. Purpose refers to motives and intents; thus, can be the designer's intention to make an automated system.

Lee and Moray (1992) explored how trust affected an operator's control strategies between human manual control and automatic machine control by conducting an experiment with a semi-automatic pasteurization plant. For this experiment, 19 subjects operated the plant with the occurrence of fault in 60 trials of 6 hours for 3 days including 10 training trials. The subjects clearly understood that safety and performance were the main goals. The subjective rating scale, developed by Muir (1989), measured the operator's trust after every trial. They found that when the operators became familiar with the plant system, performance and trust increased. When a fault occurred, performance and trust decreased; then, later recovered by the operator's adaptation over time. By comparing performance and trust, the magnitude of fault seriously affected trust because trust took more time to recover than performance. Lee and Moray (1992) suggested that

trust is determined by the overall system performance and the degree of the system component affecting operations. Findings showed the operator's control strategies can also be determined by trust in automation and self-confidence in controlling the system manually.

Trust is a crucial factor to affect the use of automation, which are human operators' control strategies. If trust is not appropriate, it can lead to misuse or disuse. Well calibrated trust can lead to appropriate use of automation. Muir (1987) mentions the importance of calibrating human trust in a machine with the use of automation to achieve good performance. According to Muir, calibration can be enhanced by improving the perception of trust, modifying the criteria of trust, boosting the user's ability to assign functions, and identify and solve poor calibration problems.

Seong and Bisantz (2008) extensively investigated calibration of trust on automation relate to decision-aid based on Sheridan (19988)'s dimensions of trust, which are reliability, robustness, validity, transparency, understandability, usefulness, and utility. Reliability and validity are related to consistent and expected function (Sheridan, 1988). Robustness is an important factor, so a decision-aid without robustness is not useful when human operators make judgement and decision (Seong & Bisantz, 2008). Transparency and understandability are related to how much human operators understand automated system. Decision-aids can improve transparency and understandability, which decrease uncertainty and increase acceptance (Seong & Bisantz, 2008). Usefulness and utility are related to effectiveness or benefit of data for decision-making, but they can be subjective to evaluate.

Appropriate trust can be defined by calibration, resolution and specificity (Lee & See, 2004). Calibration is needed between trust in automation and the automation's capabilities (Lee & Moray, 1994; Muir, 1987). Poor calibrations can happen in the situations of overtrust leading to misuse or mistrust leasing to disuse (Lee & See, 2004). When human operators overtrust automation, their level of trust is higher than actual capabilities of automation. When human operators distrust automation, their level of trust is lower than actual capabilities of automation. Good calibration means that the level of trust is corresponding with actual capabilities of automation (Lee & See, 2004). Resolution is related to correspondence between a range of trust and a range of system capabilities (Cohen et al., 1999). Specificity is related to association between the degree of trust and specific functions/entire system (Lee & See, 2004). For appropriating trust in the use of automation, good calibration, high resolution and high specificity are highly recommended. In addition, Lee and See (2004) developed a conceptual model of dynamic interaction that presents trust and its effect on reliance. This research shows how trust influences interaction with automation and how automation influences trust in a close-loop process that considers the appropriateness of trust (calibration, resolution, temporal specificity, and functional specificity), operator's individual characteristic to work (workload, engagement, self-confidence), and the cognitive process (time constraints and configuration errors).

Trust is a significant factor that influences the use of automation, because overtrust can give rise to misuse and mistrust can give rise to disuse. Therefore, it is important to calibrate appropriate trust for the most effective use of automation. With considering Sheridan(1988)'s various dimensions of trust, which are reliability,

robustness, validity, transparency, understandability, usefulness, and utility, well-designed decision-aids can assist to calibrate appropriate trust for proper use of automation (Seong & Bisantz, 2008). Also, an appropriate level of trust in automation can be defined by adjusting calibration, resolution and specificity (Lee & See, 2004).

2.4 The Human Brain Regions and EEG

Neuroimaging research are useful to investigate neural activities with a wide range of cognitive processes with attention, language and memory. Particularly, neuroimaging techniques contributes to the study of cognition. Neuroimaging techniques can assist to examine various cognitive processes including language, reasoning, problem solving, decision making and other complicated mental functions (Burge & Kahn, 2009). Neuroimaging techniques allow to find out neural activities relate to any particular cognitive task through detecting brain regions and brain waves. This part will investigate the cognitive process by examining the human brain, EEG history, EEG analysis, brain waves, brain regions, and brain imaging methodologies. Also, a clear connection to measure human trust using EEG will be articulated.

The human brain is the center of the central nervous system, which consists of the cerebrum, brainstem and cerebellum. The cerebral cortex is the outer layer of the neural tissue in the cerebrum that plays an important role in memory, attention, perception, awareness, thought, language and consciousness. The cerebral cortex is divided into four lobes – the frontal, parietal, temporal and occipital lobes. The frontal lobe involves with executive functions, the parietal lobe involves with sensory information, the temporal lobe involves with processing sensory input, and the occipital lobe involves with visual information through the visual cortex.

The cerebral cortex is relative to the cognitive process, namely, frontal lobe, parietal lobe, occipital lobe and temporal lobe (see Figure 2). The frontal lobe plays a crucial role in the brain as a control tower because of its association with memory, willpower, logical thinking, adjusting the activities of other regions of the brain, and controlling emotion and impulse (Metcalf, 2016). In addition, the frontal lobe is associated with problem-solving, judgment, attention, organization, planning, and anticipation (Fuster, 2002). Particularly, the prefrontal cortex (PFC), which is the cerebral cortex that covers the front part of the frontal lobe, performs executive functions. This involves organizing and executing a series of actions for complex cognitive goals by predicting consequences of current activities, and making decisions by differentiating among conflicting thoughts and determining good and bad (Yang & Raine, 2009; Fuster, 2002). To conduct actions in a timely manner, the prefrontal cortex needs direct access to sensory, motor and mnemonic information (Fuster, 2002). A damaged prefrontal cortex could result in antisocial, violent and psychopathic behaviors (Yang & Raine, 2009).

The parietal lobe of the brain is located above the occipital lobe and behind the frontal lobe and central sulcus. The parietal lobe deals with language, sensation, perception and integrating sensory inputs, especially the sense of touch, pain and temperature. The posterior parietal cortex is highly associated with visuomotor integration, spatial perception and spatial orientation, because of its location between the sensory and motor area that are related to both sensory-related and motor-related activity (Buneo & Andersen, 2006; Kolb & Whishaw, 2011).

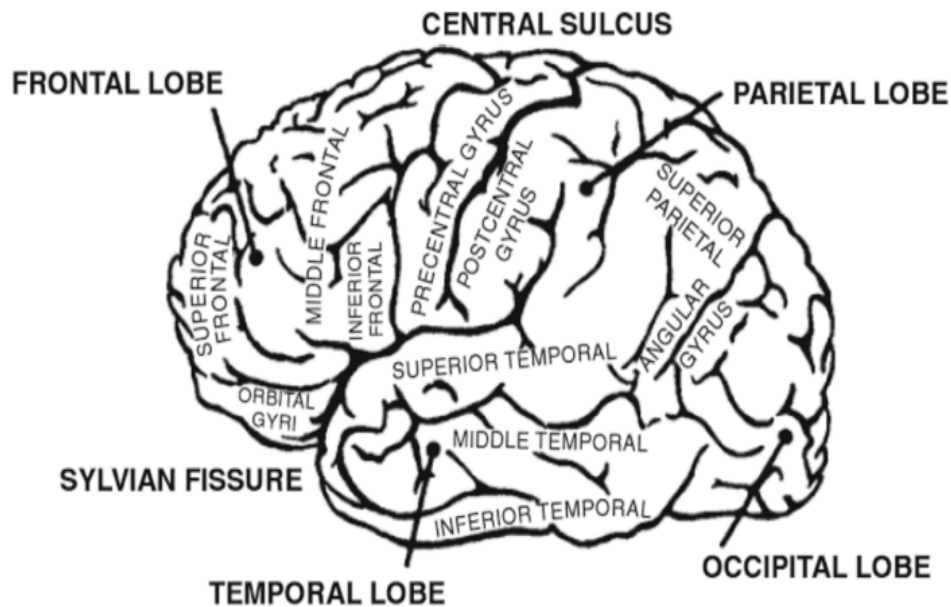


Figure 2: Brain Anatomy (Semrud-Clikeman & Teeter 2011)

The occipital lobe is located under the occipital bone at the lower back of the skull and is divided into two areas, dorsal (superior) and ventral (inferior). As a visual processing center, it houses the visual cortex which functions visuospatial processing, color differentiation, and motion perception (Kolb & Whishaw, 2003). Damage to the right hemisphere of the occipital lobe can cause serious problems with the visual perception (Semrud-Clikeman & Teeter, 2009).

The temporal lobe is interconnected with other cortical regions for executing complicated functions (Semrud-Clikeman & Teeter, 2009). This lobe is associated with the perception of visual and auditory sensation related to speech and language processing (Read, 1981), language comprehension, long term memory storage and facial and object recognition (Bauer & Demery, 2003) and emotion. A damaged temporal lobe can cause perseverative speech, paranoia, and rage (Benson & Blumer, 1975). Amygdala, which is deep inside the temporal lobe is known for controlling major affective activities like friendship, love and affection and the expression of moods such as fear, rage, and aggression (Phelps, 2004). The amygdala is the center for identifying danger that aids in self-protection, when activated brings fear and anxiety (Phelps, 2004).

The human brain functions when neurons, electrically excitable cells process and transmit information, are electrically charged by membranes that transport proteins to pump ions across their membranes and pass electrical signals to other neurons through a synapse, a specialized connection with other cells. Neurons make up the core components of the brain and the spinal cord (the central nervous system (CNS)), and the ganglia (the peripheral nervous system (PNS)). These electrical signals affect all the processes in the brain; therefore, understanding how the brain works is important (e.g., how neurons communication, how stimuli are maintained to function correctly, how attention is used

selectively, and how memory is made and stored to understand how neurons communicate). These electrical signals can be detected and recorded using electroencephalography (EEG). EEG is an electrophysiological monitoring method used to record the electrical activity of the brain (Evans & Abarbanel, 2009). EEG measures voltage fluctuations that resulted from ionic current within the neurons of the brain detected by multiple electrodes placed on the scalp.

The use of EEGs to measure brain activity dates back more than a century ago. In 1875, Richard Caton (1842-1926), an English physician, reported electrical activities from the cerebral cortex of rabbits and monkeys in the *British Medical Journal*. German physiologist and psychiatrist Hans Berger (1873-1941) recorded the first electrical activities of human brain by inserting two platinum electrodes under the damaged cranial bone and called it electroencephalography (EEG) similar to an electrocardiogram (ECG) or electromyography (EMG).

2.5 EEG Analyses (Wavelet and Fourier Analysis/Power Spectrum)

EEG provides a visual representation of brain waves as complex patterns of vibrating electrical signals; therefore, there is no value in analyzing the complex patterns directly. A digital signal analysis of EEG includes two analyses, Fourier and Wavelet.

The Fourier analysis decomposes the EEG time series into a voltage-by-frequency spectral graph, also called power spectrum, is the most common EEG analysis. For clarity, this research refers to Power Spectrum. Power spectrum assumes the brain waves are linear combinations of simple oscillators that vibrate at specific frequencies, so power is the square of the EEG magnitude and magnitude is the integral average of the amplitude of the EEG signal measured from + peak to – peak across the time sample. The wavelet analysis is made to supplement the traditional Fourier analysis. It is a wave-like oscillation with an amplitude that begins at zero, increases, and then decreases back to zero. It is a time-frequency technique that transforms EEG signals over time.

2.6 EEG Band related to Cognitive Process

Brain waves consist of different frequency waves, which allows one to analyze the waves using a specific frequency. The Power Spectrum analysis allows for classifying the brain waves by frequency ranges: delta waves (0.2-4 Hz), theta waves (4-8 Hz), alpha waves (8-13 Hz), beta waves (13-30 Hz) and gamma waves (30-60 Hz). The classification of the beta waves is low beta (13-15 Hz, which is called as SMR), mid beta (15-20 Hz) and high beta (20-30 Hz) (Sanei & Chambers, 2007; Pfurtschella, Lopes, & Silva, 1999).

Delta waves have a frequency between 0.2 Hz and 4 Hz. Usually, these waves are common in infants or in normal adults in a deep state of sleep. They tend to be the highest in amplitude and the slowest waves. When delta waves are strong, the body is healing itself by resetting its internal clock (Botella-Soler, Valderrama, Crépon, Navarro, & Le Van Quyen, 2012). Delta waves are common in people in a coma or have a defect of the cerebral cortex. Delta waves frequency and the frequency of the movement of the eyes or bodies to objects are similar. During an experiment conducted to record brain waves over time due to unavoidable noise, the increased power of the delta waves was not seriously considered for analysis.

Alpha waves have a frequency between 8 Hz and 13 Hz. The waves appear when people are awake but have a relaxed awareness and inattention without processing much information. When people are more comfortable and relaxed, the alpha waves increase. The waves generally appear as a regular wave, detected strongly over the occipital lobe, and in small amounts over the frontal lobe. They emerge with the closing of the eyes and with tension, and are suppressed during eye opening or mental exertion. Cahn and Polich (2006) detected strong alpha activity in experienced meditators. These waves relate to recalling memory, relieving pain and reducing stress and anxiety (Williams, Ramaswamy, & Oulhaj, 2006; Williams, 2001; Nomura, Higuchi, & Yu, 2006; Ossebaard, 2000). Therefore, stimulating alpha waves can improve the efficiency of learning and working.

Beta waves have a frequency between 13 Hz and 30 Hz. These waves are detected mostly in the front lobe and are associated with conscious activities. They can affect cognitive tasks such as attention, problem solving, judgment and decision-making. In sustained attention tasks, low beta (SMR) waves affect perceptual sensitivity and lower omission errors, while mid-beta waves reduce reaction times (Egner & Gruzelier, 2004; Lqbai, Muhammed, Khan, & Farooq, 2016). Beta waves can be a guideline for measuring the cognitive process (Friel, 2007). High beta waves are associated with intensity, anxiety and hypervigilance (Friel, 2007). People that lack sufficient beta activity can experience mental or emotional disorders such as depression, Attention Deficit Disorder (ADD) (Brenner, Ulrich, Spiker, Sciabassi, Reynolds III, Marin, & Boller, 1986; Egner & Gruzelier, 2004) and insomnia. Beta activity is important because it can enhance concentration, attention, emotional stability and energy levels (Siever, 2004; CE, LE, & SJ, 1986; GJ, 1996).

Gamma waves have a frequency between 30 Hz and 60 Hz. These are the fastest waves, indicative of anxiety and processes advanced cognitive information such as reasoning and judgement. They are connected to the development of ideas, the use of language, processing memories and learning (Crone, Hao, Hart, Boatman, Lesser, Irizarry, & Gordon, 2001; Burle & Bonnet, 2000; Miltner, Braun, Arnold, Witte, & Taub, 1999). They are detectable when people use their short-term memory to identify objects, sounds and tactile sensations.

Theta waves have a frequency between 4 Hz and 8 Hz. These waves occur in young children and are present in adults and teens during drowsiness conditions. They are usually detected when people are falling asleep or a calming down state. Theta waves are associated with access to unconscious mind, creative inspiration and deep meditation, but this has not been proved because of lack of standardized results. They can be used in hypnotherapy and self-hypnosis during a receptive mental state to accept affirmations and suggestions (Wickramasekera, 1977; Sabourin, Cutcomb, Crawford, & Pribram, 1990).

2.7 Comparisons of EEG vs CT, MRI, fMRI, PET and MEG

There are two methods to measure brain activities, which are invasive and non-invasive. The invasive method requires physical surgeries, which are used in the experiments of monkeys or brain patients. The invasive method measures brain signals inside the skull directly. Therefore, this method which measures signal strength relative to background noise can acquire a high signal-to-noise ratio (SNR). The non-invasive

method can be used for a wide range of experiments that does not require surgery. Because this method measures brain signals outside the skull it can lose any important information from weak or sensitive signals and include background noise. The two methods described here are brain imaging techniques such as CT, MRI, fMRI, PET, MEG and EEG (Dale, Liu, Fisch, Buckner, Belliveau, Lwwine, & Halgren, 2000)

Computerized Tomography (CT) is a medical imaging technique that uses computer-processed combinations of many X-ray images taken from different angles to produce cross-sectional images of specific areas of a scanned object, allowing the user to see inside the object without cutting. (Mitsopoulos, Baumgartner, Heymsfield, Lyons, Gallagher, & Ross, 1998). CT can provide three-dimensional imagery of the brain and accurately diagnoses medical conditions with improved resolution. However, CTs are known to the risk of increase cancer as a result of x-ray radiation exposure and possible allergic reaction can occur from the injection of intravenous agents. (Kobayshi, Shimoda, Nakagawa, & Yamamoto, 2004).

Magnetic resonance imaging (MRI) is a medical imaging technique to create pictures of the inside the body with using a nuclear magnetic resonance of protons to produce proton density images (Huettel, Song, McCarthy, 2004). When hydrogen atoms generate a detectable radio-frequency signal, the signal is received by antennas in close proximity to the anatomy being examined. MRI helps examine tissues and fluid in addition to bone and provides cross two-dimensional imaging such as cross section, longitudinal section and side section through body parts so physicians can have a three-dimensional of the anatomy by combining the images (Huettel, Song, McCarthy, 2004). The principle of MRI is that when hydrogen atoms are activated by radio frequency waves in a magnetic field, the hydrogen atoms are emitting (Pinel, 2011). MRI generates high spatial resolution, showing different spatial locations and three-dimensional images from measuring the waves in the hydrogen atoms emitting process. The high spatial resolution images are much more explicit than CT (Kim, Kim, Yun, Sohn, & Min, 1999).

Functional magnetic resonance imaging (fMRI) is a newly developed method for studying the brain function using blood flow to the functioning areas of the brain. When the brain region is active, it requires more oxygen, which is delivered by increasing the blood flow (Matthews, Jezzard, 2004). fMRI allows scientists to detect the active part of the central nervous system's response to the assigned task by measuring blood oxygen levels in the brain. There is a difference in the blood flow between the active area and the inactive or resting area of the brain; therefore, scientists can compare differences between assigned tasks. fMRI provides accurate structural data and a reasonable temporal resolution. MRI and fMRI provides excellent spatial resolution and good temporal resolution without no radiation and no side effects from the magnetic fields and radio waves. However, both MRI and fMRI are claustrophobic, noisy, and very sensitive to moving objects. MRI and fMRI requires special managing because metal-based equipment cannot be around the machine. Also, any presence of metal in the body can be a safety hazard in both MRI and fMRI (Huettel, Song, McCarthy, 2004).

Positron Emission Tomography (PET) is a nuclear medicine and functional imaging technique that produces a three-dimensional image or picture of functional processes in the body (Gambhir, 2002). When a positron-emitting radionuclide is inserted into the body in a biologically active molecule, the PET system can detect pairs of

gamma rays emitted by it, which is known as a tracer. The images from the tracer concentration are developed into three- or four-dimensional images after time spent in the body, and the images are reconstructed by computer analysis. PET can be used with CT X-ray or fMRI for better diagnosis or experimentation. There are several advantages of using PET. 1) PET can provide metabolic activity within the brain, so neurologists can understand neural functionality, structural abnormalities, and biochemical change such as neurotransmission. 2) PET can diagnose an array of neurological disorders and conditions by detecting irregularities in the body function caused by disease generated in the brain that shows different levels of radioactivity. 3) PET also is not claustrophobic, and there is no problem with motion artifacts. Disadvantages of PET are 1) PET does not provide temporal resolution and structural information, and 2) has poor spatial resolution (Huettel, Song, McCarthy, 2004); 3) PET's radioactive tracers may provoke allergic reactions and lead to adverse long-term effects; and 4) PET is expensive and not portable.

Magnetoencephalography (MEG) is a functional neuroimaging technique. This technique maps brain activity by recording magnetic fields generated by electrical current that occur naturally in the brain. This is done by using very sensitive magnetometers (Baumgartner & Pataria, 2006). MEG determines changes in the magnetic fields on the surface of the scalp. MEG is different from fMRI, because the magnetic field occurs as a result of changes in neural activity, which activate pyramidal cells of the cortex; the latter are not affected by magnetic fields. MEG is used to research the perceptual and cognitive brain processes, for localizing the regions affected by pathology before surgical removal, for determining the function of different regions of the brain and for neurofeedback (Whelless, Castillo, Maggio, Kim, Breier, Simos, & Papanicolaou, 2004). MEG provides high temporal resolution, acceptable spatial resolution and less distortion by the scalp compared to EEG. However, MEG requires high baselines from the firing rate, it is expensive, and is not portable.

Electroencephalography (EEG) is an electrophysiological monitoring method that records the electrical activity of the brain by placing electrodes on the scalp. EEG has several benefits for measuring brain activity. EEG detects changes in the brain's electrical activity in various regions within a millisecond timeframe in real time. Therefore, EEG can generate an accurate result compared to PET and fMRI, which have a time resolution of between seconds and minutes. A single neuron generally takes approximately 0.5-130 milliseconds to communicate with another neuron. Therefore, the millisecond timeframe of EEG is more appropriate for recording brain activity in real time (Munari, Hoffmann, Fracione, Kahane, Tassi, Lo Russo, & Benabid, 1994). EEG directly measures the electrical activity of the brain with electrodes on the scalp; whereas, other techniques indirectly record brain activity in the blood flow or metabolic activity (Debener, Ullsperger, Siegel, & Engel, 2006). EEG provides high resolution without exposure to radiation or magnetic fields. In addition, EEG is not expensive compared to other equipment types, portable, and is not claustrophobic. Most importantly, EEG can detect the response to the external stimuli through recording the brain's electrical activity in a real time, so it is useful for scientific experiments with a wide range of cognitive tasks. By considering all neuroimaging techniques, I propose to measure and analyze human brainwaves using situations involving trust and mistrust to understand the cognitive process using EEG.

CHAPTER 3. METHOD

In order to understand the cognitive process, this research will measure and analyze human brainwaves in situations involving trust and mistrust and by using an electroencephalogram (EEG). The overall purpose of the proposed study is to investigate the human cognitive process of trust and mistrust using EEG. More specifically, the proposed study seeks to conduct experiments of two phases: Phase I and Phase II. The phase I experiment will use a word elicitation study to identify specific brainwaves and brain regions resulting from situations of trust and mistrust. By identifying specific brainwaves (i.e. delta, theta, alpha, beta and gamma) and brain regions (i.e. frontal, parietal, temporal and occipital lobes), Phase I can enable to discover cognitive functions and activities relate to trust and mistrust. The Phase II experiment will use a simulation to evaluate strategies of individual human operator's use of automation in the situations of trust and mistrust. The simulate driving game is designed to observe participant's use of automation when trust in automatic control is developed in the first five trials and when mistrust in automatic control is developed in the second five trials. Through this simulation, it is expected to evaluate human operator's use of automation in trust and mistrust. Eventually this proposed research will investigate the cognitive processes of trust and mistrust by analyzing corresponding data among identified brainwaves, stimulated brain regions, and the strategies of individual human operator's use of automation.

3.1 Hypotheses

For all of the reason mentioned above, a further in-depth investigation into human cognitive processes is needed in order to understand human trust on automation. The proposed study will investigate the human cognitive process related to trust and mistrust through EEG recording, identify and analyze specific brainwaves and brain regions, and explore how the cognitive process of trust and mistrust affect human operator's use of automation. Based on the findings of current literature and the rationale describe, the followings are hypotheses:

- In the situation of trust, specific brainwaves will be detected, which will correlate to normal cognitive process with a calm and relaxed state.
- In the situation of trust, specific brain regions will be activated, which will correlate normal cognitive process with executive functions.
- In the situation of mistrust, specific brainwaves will be detected, which will correlate to cognitive processes with significant stress and anxiety.
- In the situation of mistrust, specific brain regions will be detected, which will correlate to intense emotions.
- When a human operator trusts automation, specific brainwaves and brain regions will be detected, and human operator's strategy the use of automation will increase.
- When a human operator mistrusts automation, specific brainwaves and brain regions will be detected, and human operator's strategy on the use of automation will decrease.

3.2 Participant

A total of 20 participants (male and female) would be solicited to participate in the proposed experiment. All participants would encompass undergraduate and graduate students recruited on campus. Each participant will be over 18 years old, with normal or corrected to normal vision, will be free of current or past neurological and psychiatric disorders, and will be able to read and comprehend the English language. There will be no preference on right-handed or left-handed user, but all participants must have the ability to use a keyboard with their hands. There will be no discrimination upon participants based on age, race, gender, religion, or prior participation in other experiments. Each participant will be in an isolated room under guided instruction. Conditions for all participants will be the same.

3.3 Instrumentation

The EEG data will be recorded using the g.HIamp (256 multichannel amplifier), g. GAMMASys (active electrode system with g. GAMMAcap) and g. Recorder (brain signal recording software) by g. tec medical engineering company (see Figure 3). Per the International 10-20 system of electrode placement, 20 electrodes (Fp1, Fp2, Fpz, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, T6, O1, O2) are recorded and measured (see Figure 4). The 10-20 system is based on the relationship between the location of an electrode and the underlying area of the cerebral cortex. The letters F, T, C, P, and O represent the frontal, temporal, central, parietal, and occipital lobes. Even numbers refer to the right hemisphere of the brain and odd numbers refer to the left hemisphere of the brain. The participant's head is fitted with a cap of electrodes (g. GAMMAcap), and the selected 20 electrodes are filled with abrasive electrolyte gel using a syringe. The sampling frequency to record brainwaves is 256 Hz, high-frequency filter as 50 Hz, and low-frequency filter as -50 Hz.

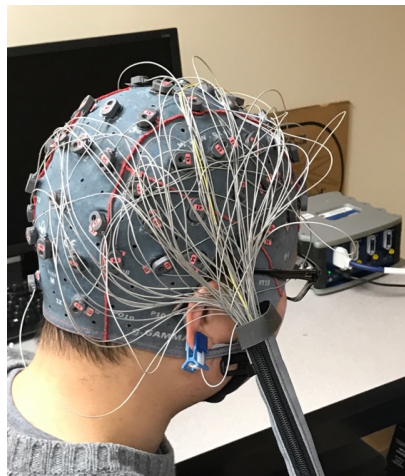


Figure 3: Example of g. GAMMAcap & g.HIamp

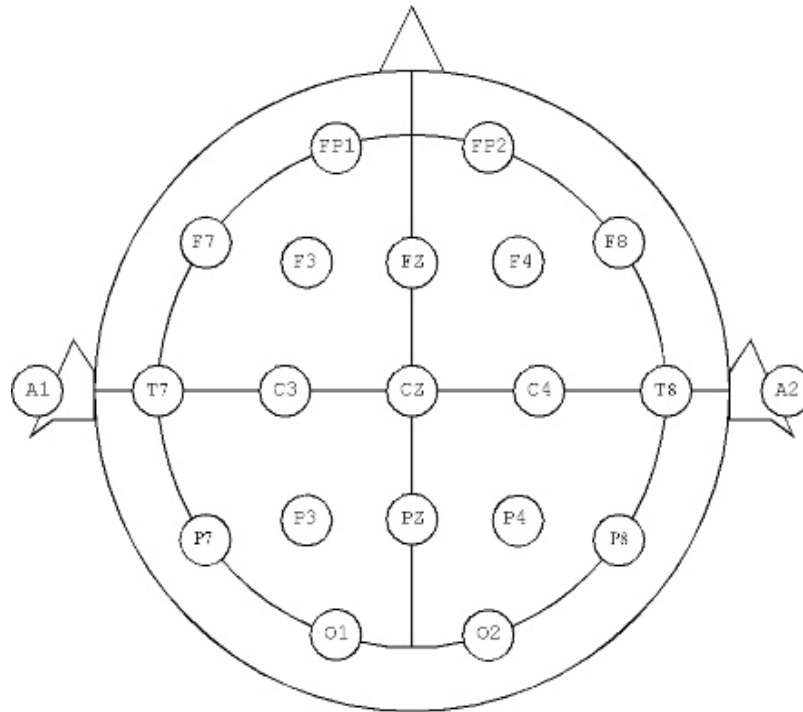


Figure 4: International 10-20 system for electrode placement

3.4 Experiments

The overall experiment consists of two phases: Phase I and Phase II. The following sections will provide details regarding the experiment associated with each phase.

3.4.1. Procedure

Individual participants will be instructed to read an informed consent form. Any question and concerns will be addressed. The participant will be required to provide demographic information such as age, gender, and contact information as well as a signed and dated consent form. After each participant signs the consent form, the participant will be asked to put aside all personal items on a secure table. The participant will be asked to turn off their cell phone to avoid any interruption during the experiment. After a participant takes a seat, the instruction for each experiment will be presented in a detail. Any question or concern related to the experiment will be addressed and answered by a facilitator in the room. The participant will be fitted with a g. GAMMA cap with electrodes that will measure small electrical signals produced by the brain in a completely non-invasive manner. The 20 electrodes, per the International 10-20 system for electrode placement, will be filled with a gel using a syringe with a plastic tip. The participant will be asked to focus their attention on the computer screen. After the experiment, each participant will gently take off the electrode cap and remove any remaining gel from their head or hair using provided tissues and/or towel.

3.4.1 Phase I: experiment with words

This study adopts the framework from a previous study that asks participants to rank words related to trust and mistrust. This study utilizes only the first 15 ranked words related to trust and mistrust (Jian, Jiun-Yin, Bisantz, & Drury, (2000); Bistantz, Finger, Seong, & Llians (1999). The participants will be expected to think about situations of trust and mistrust evoked by the selected words. They will be asked not to talk and blink their eyes (if possible) to reduce noise in recording brainwaves.

All 20 subjects will participate in the experiment with words. After fitting each participant with a cap of electrodes, they will sit in front of the computer screen. 10 seconds before the experiment begins the screen will display instructions. Each word displayed will last for 7 seconds in a slideshow, which is enough time to read and understand the meaning of each word. The slideshow will last for a total of 3 minutes and 30 seconds, which is enough time to evoke a situation of trust or mistrust. Each experiment will consist of two sessions: 1) trust words and 2) mistrust words.

The first session will consist of a slideshow displaying 30 random stimuli from the 15 words related to trust, which will be enough stimuli for recording brainwaves (see Figure 5). The EEG will record the participant's brainwaves for trust. The participant will take a one-minute break to avoid fatigue and stress. Similarly, the second slideshow will consist of 30 random stimuli from the 15 words related to mistrust, which will be enough stimuli for recording brainwaves (see Table 1). The EEG will record the participant's brainwaves for mistrust.

Table 1. Words lists of trust and mistrust for study 2

	Words related to Trust	Words related to Mistrust
1	Honesty	Cheat
2	Loyalty	Betray
3	Love	Deception
4	Confidence	Steal
5	Assurance	Suspicion
6	Friendship	Distrust
7	Security	Sneaky
8	Integrity	Misleading
9	Fidelity	Mistrust
10	Familiarity	Phone
11	Honor	Beware
12	Reliability	Harm
13	Trustworthy	Falsity
14	Entrust	Lie
15	Promise	Cruel

3.4.2 Phase II: experiment with simulation

The experiment will require each participant to become a human operator that controls a simulated driving game. For each participant, a PC computer will provide a simulated driving experience in the format of an interactive game using a keyboard. This simulated driving game will be designed to have 10 obstructions on the road within a two-minute driving period. At the start of each simulated driving period, each participant must choose between an automatic or a manual control of the movement of the vehicle. All 20 participants will individually participate in 10 trials of the simulated driving experiment.

As the operator, each participant will be given a goal of performance by avoiding accidents. All participants will be provided incentives (yet to be determined) to perform well. Performance will be measured by the number of accidents occurred during each driving trial. Successfully avoiding all 10 obstacles would result in a 100% performance rate, and running into all 10 obstacles would result in a 0% performance rate. A complete reliance on automatic control will have a 90% performance rate or less.

Before operating the simulated vehicle, each participant will receive written descriptions that inform the objective of the game. This description will be delivered via a video that demonstrates the controls of the game (manual or automatic, using specific keys on the keyboard, etc.) while also explaining the possibilities of running into obstacles. The demonstration video will also show two kinds of driving experience—one using only automatic control and the other using only manual control, and an example of running into an obstacle for each driving mode. Any question or concern related to the objective and operation of the game will be addressed and answered by a facilitator in the room. After the video demonstration, each participant will be fitted with a EEG cap that is mounted with electrodes. The participant will sit in front of the computer screen with a keyboard and begin experimental trials.

All participants will individually operate a moving vehicle in a driving simulation through 10 trials, each trial lasting two minutes (See Table 2). Using assigned keyboard keys, each participant will begin operating by choosing either an automated or manual control at the beginning for each trial. During each trial, the screen will display the current position of the moving vehicle, whether automated or manual control was chosen, and the number of accidents occurred (ranging from 0 to 10) (See Figure 5). The position of obstacles will randomly change from one trial to the next.

Trials 1 to 5 (which would last a total of 10 minutes) will be designed for the participant to develop trust in automatic control by showing consistent and reliable performance (at the rate of 90% resulting from 1 accident). Trials 4 and 5 will be designed for an expected formation of trust, during which each participant's brainwaves and brain regions will be recorded and analyzed (by considering the participant's decision between automated or manual control for each trial). Trials 6 through 10 (which would last an additional 10 minutes) will be designed for the participant to develop mistrust in automatic control by providing an increasing number of accidents for each successive trial. Trials 9 and 10 will be designed to have an expected formation of mistrust, during which each participant's brainwaves and brain regions will be recorded and analyzed (by

considering the participant's decision between automated or manual control for each trial). In order to avoid interruption and noise in recording brainwaves and brain regions, there will be no break in all 10 trials. A more detailed experiment plan will be further developed.

Table 2: Plan for Experiment

Driving Trial	Demo Video Auto	Demo Video Manual	Break time	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9	Trial 10
Number of Accidents via Automatic Control	1	1		1	1	1	1	1	3	4	5	5	5
EEG recording				EEG	EEG	EEG	EEG	EEG	EEG	EEG	EEG	EEG	EEG
Expected Trust and Mistrust Situations							Trust	Trust				Mis-trust	Mis-trust



Figure 5: Example of the Simulated Driving

3.5 Data Collection and Analysis

This research used both power spectrum for identifying specific brainwaves and coherence analysis for detecting active brain regions. Analyzing brainwaves consisted of linear spectral measurements, nonlinear measurements of complexity, and interdependency measurements (Teplan, 2006). The benefits of using linear spectral measurements (power spectrum and coherence analysis) are results of obtainable fast and easy; thus, most often used.

One linear spectral measurement termed the power spectrum measurement is a common method to quantify EEG raw data to illustrate the distribution of signal power by frequency (Teplan, Krakovska, & Stolc, 2006). To identify specific brainwaves corresponding to the words for trust and mistrust situations, this involved analyzing the EEG data using the power spectrum through the EMSE Suite Data Editor by Cortech Solutions. Power spectrum analysis involved performing a Fast Fourier Transform (FFT)

on defined data intervals. Then, the results are squared and averaged. The power spectrum data is displayed in units of Volts².

$$1.0 \text{ uV}^2 = 1.0 (\text{uV})^2 = 1.0\text{e-}012 \text{ V}^2. \quad (1)$$

Power spectrum analysis is helpful in analyzing each channel's frequency; however, this analysis has limitations when examining complex and comprehensive activities of the brain such as trust and mistrust situations. Therefore, this study used a coherence analysis to detect active brain regions and to examine communications among the active brain regions.

Using coherence analysis (based on the power spectrum) for functional connectivity in the different brain regions has gained frequent use (Thatcher, 1986). Coherence is a measurement of the amount of phase stability or phase jitter between two different time series. When the phase difference between two signals is constant, coherence = 1; when the phase difference between signals is random, coherence = 0. To obtain coherence when two channels (a, b) are recorded from a specific range of the frequency (f); the square of cross-spectrum (P_{ab}) needs to be divided by power spectrum $P_a(f)$ and multiplied by power spectrum $P_b(f)$.

$$Coh_{ab}(f) = \frac{P_{ab}(f)^2}{[P_a(f) \times P_b(f)]} \quad (2)$$

References

1. Avanzo, C. D., Tarantino, V., Bisiacchi, P., & Sparacino, G. (2009). A wavelet methodology for EEG time-frequency analysis in a time discrimination task. *International Journal of Bioelectromagnetism*, 11(4), 185-188.
2. Barber, B. (1983). *The Logic and Limits of Trust*. New Brunswick. Rutgers University Press.
3. Baumgartner, Christopher, & Patariaia, Ekaterina (2006). Revisiting the role of magnetoencephalography in epilepsy, *Current opinion in neurology*.
4. Benson, D. F., & Blumer, D. (1975). Personality changes with frontal and temporal lobe lesions. *Psychiatric aspects of neurological diseases*.
5. Bhattacharya, R., Devinney, T. M., & Pillutla, M. M. (1998). A formal model of trust based on outcomes. *Academy of management review*, 23(3), 459-472.
6. Bisantz, A. M., & Seong, Y. (2001). Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics*, 28(2), 85-97.
7. Bisantz, A., Llinas, J., Seong, Y., Finger, R., & Jian, J. Y. (2000). Empirical Investigations of Trust-Related Systems Vulnerabilities in Aided, Adversarial Decision Making. *STATE UNIV OF NEW YORK AT BUFFALO CENTER OF MULTISOURCE INFORMATION FUSION*.
8. Botella-Soler, V., Valderrama, M., Crépon, B., Navarro, V., & Le Van Quyen, M. (2012). Large-scale cortical dynamics of sleep slow waves. *PloS one*, 7(2).
9. Brenner, R. P., Ulrich, R. F., Spiker, D. G., Sciabassi, R. J., Reynolds, III, C. F., Marin, R. S., & Boller, F. (1986). Computerized EEG spectral analysis in elderly normal, demented and depressed subjects. *Electroencephalography and clinical neurophysiology*, 64(6), 483-492.

10. Budzynski, Thomas H., Budzynski, Helen Kogan, Evans, James R., & Abarbanel, Andrew (2009). Introduction to Quantitative EEG and Neurofeedback
11. Buneo, Christopher A., & Anderson, Richard A. (2006). The Posterior Parietal Cortex: Sensorimotor Interface for the Planning and Online Control of Visually Guided Movements. *Neuropsychologia*, 2594-2606.
12. Bunge, S. A., & Kahn, I. (2009). Cognition: An overview of neuroimaging techniques. *Encyclopedia of neuroscience*, 2, 1063-1067.
13. Burle, B., & Bonnet, M. (2000). High-speed memory scanning: a behavioral argument for a serial oscillatory model. *Cognitive Brain Research*, 9(3), 327-337.
14. Cahn, B. R., Polich, J. (2006). Meditation states and traits: EEG, ERP, and neuroimaging studies. *Psychol Bull.* 132(2), 180-211.
15. Cohen, M. S., Parasuraman, R., & Freeman, J. T. (1998). Trust in decision aids: A model and its training implications. In *Proc. Command and Control Research and Technology Symp.*
16. Crone, N. E., Hao, L., Hart, J., Boatman, D., Lesser, R. P., Irizarry, R., & Gordon, B. (2001). Electrographic gamma activity during word production in spoken and sign language. *Neurology*, 57(11), 2045-2053.
17. Dale, Anders M., Liu, Arthur K., Fischl, Bruce R., Buckner, Randy L., Belliveau, John W., Lewine, Jeffrey D., & Halgren, Eric (2000). Dynamic Statistical Parametric Mapping: Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity. *Neuron*, 26(1), 55-67.
18. Debener, Stefan, Ullsperger, Markus, Siegel, Markus, & Engel, Andreas K. (2006). Single-trial EEG-fMRI reveals the dynamics of cognitive function. *Trends in Cognitive Science*, 10(12), 558-563.
19. Deutsch, M. (1973). The Resolution of Conflict: Constructive and Destructive Processes. *New Haven, CN: Yale University Press.*
20. Dimoka, Angelika (2011). Brain mapping of psychological processes with psychometric scales: An fMRI method for social neuroscience. *NeuroImage*, 54(1), S263-S271.
21. Egner, T., & Gruzelier, J. H. (2004). EEG biofeedback of low beta band components: frequency-specific effects on variables of attention and event-related brain potentials. *Clinical Neurophysiology*, 115(1), 131-139.
22. Egner, T., & Gruzelier, J. H. (2004). EEG biofeedback of low beta band components: Frequency-specific effects on variables of attention and event-related brain potentials. *Clinical Neurophysiology*, 115(1), 131-139.
23. Fahey, L. J. (2007). *Arleigh burke grounding INS plot*. Unpublished manuscript.
24. Farber, E., & Paley, M. (1993). Using freeway traffic data to estimate the effectiveness of rear-end collision countermeasures. *Paper presented at the Third Annual IVHS America Meeting, Washington, DC* (April).
25. Friel, Patrick N. (2007). EEG Biofeedback in the Treatment of Attention Deficit/ Hyperactivity Disorder. *Alternative Medicine Review*, 12(2).
26. Fuster, Joaquin M. (2002). Frontal lobe and cognitive development. *Journal of Neurocytology*, 31, 373-385.
27. Gambhir, S. S. (2002). Molecular imaging of cancer with positron emission tomography. *Nature Reviews Cancer* 2, 683-693.
28. Howard, C. E., Graham, L. E., II, Wycoff, S. J., A comparison of methods for reducing stress among dental students. *J Dent Educ.* 1986;50(9):542-544
29. Huettel, S. A., Song, A.W., & McCarthy, G. (2004). *Functional magnetic resonance imaging second edition.*

30. Jian, Jiun-Yin, Bisantz, Ann M., & Drury, Colin, G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International journal of cognitive ergonomics*, 4(1), 53-71.
31. Kim, K. N., Kim, M. J., Yun, S. H., Sohn, S. K., & Min, J. S. (2004). Comparative study of transrectal ultrasonography, pelvic computerized tomography, and magnetic resonance imaging in preoperative staging of rectal cancer. *Diseases of the Colon & Rectum*, 42(6), 770-775.
32. Kobayashi, Kaoru, Shimoda, Shinji, Nakagawa, Yoichi, & Yamamoto, Akira (2004). Accuracy in Measurement of Distance Using Limited Cone-Beam Computerized Tomography. *International Journal of Oral & Maxillofacial Implants*, 19(2), 228-231.
33. Kolb, Bryan, & Whishaw, Ian Q. (2009). Fundamentals of Human Neuropsychology. *Basingstoke, Palgrave Macmillan*.
34. Kolb, Bryan, & Whishaw, Ian Q. (2011). An Introduction to Brain and Behavior. *New York, Worth Publ.*
35. Krueger, Frank, McCabe, Kevin, Moll, Jorge, Kriegeskorte, Nikolaus, Zahn, Roland, Strenziok, Maren., Heinecke, Armin, & Grafman, Jordan (2007). Neural correlates of trust. *PNAS*, 104(50), 20084-20089.
36. Larzelere, R. E. & Huston, T. L. (1980). The dyadic trust scale: Toward understanding interpersonal trust in close relationships. *Journal of Marriage and the Family*, 595-604.
37. Lee, J. D. & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, 40, 153- 184.
38. Lee, J. D. (2008). Review of a pivotal Human Factors article: Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors* 50(3), 404-410.
39. Lee, J. D., & Moray, N. (1994). Trust, control strategies and allocation of function in human-machine system. *Ergonomics*, 35, 1243-1270.
40. Lee, J. D., & See, K. A. (2004). Trust in technology: Designing for appropriate reliance. *Human Factors*, 46, 50–80.
41. Lerch, F. J., & Prietula, M. J. (1989). How Do We Trust Machine Advice? Designing and using human-computer interfaces and knowledge based systems. *Paper presented at the Proceedings of the Third Annual Conference on Human Computer Interaction*.
42. Lqbai, Sadaf, Muhammed, Shanir, P. P., Khan, Yusuf, Uzzaman, & Farooq, Omar (2016). EEG Analysis of Imagined Speech. *International Journal of Rough Sets and Data Analysis*, 3(2).
43. Masalonis, A. J., R. Parasuraman, et al. (1999). Trust as a construct for evaluation of automated aids: Past and future theory and research. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting, Vols 1 and 2*: 184-188.
44. Matthews, P. M., & Jezzard, P. (2004). Functional magnetic resonance imaging. *Journal of Neurology, Neurosurgery*.
45. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20, 709–734.
46. Metcalf, Brain R. (2016). Frontal Lobe Function. *Encyclopedia of Theory in Psychology Thousand Oaks*, 340-344.
47. Miltner, W. H., Braun, C., Arnold, M., Witte, H., & Taub, E. (1999). Coherence of gamma-band EEG activity as a basis for associative learning. *Nature*, 397(6718), 434-436.
48. Mitsiopoulos, N., Baumgartner, R. N., Heymsfield, S. B., Lyons, W., Gallagher, D., & Ross, D. (1998). Cadaver validation of skeletal muscle measurement by magnetic resonance imaging and computerized tomography. *Journal of Applied Physiology*, 85(1), 115-122.

49. Moorman, Christine, Rohit, Deshpande, & Gerald, Zaltman (1993). Factors Affecting Trust in Market Research Relationships. *Journal of Marketing*, 57(January), 81-101.
50. Muir, B. M. & Moray, N. (1996). Trust in automation: Part U. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39(3), 429-460.
51. Muir, B. M. (1987). Trust Between Humans and Machines, and Design of Decision Aids, *International Journal of Man-Machine Studies*, 27, 527-539.
52. Muir, B. M. (1989). Operators' trust in and use of automatic controllers in a supervisory process control task. *Unpublished Ph.D. dissertation, University of Toronto, Toronto, Canada.*
53. Munari, C., Hoffmann, D., Fracione, S., Kahane, P., Tassi, L., Lo Russo, G., & Benabid, A. L. (1994). Stereo-electroencephalography methodology: advantages and limits. *Acta Neurologica*, 89(S152), 56-67.
54. National Transportation Safety Board (1973). Eastern Airlines L-1011, Miami, Florida, 20 December 1972 (Rep. NTSB- AAR-73-14). *Washington, DC: Author.*
55. National Transportation Safety Board (1988). Railroad accident report, Maryland, 4 January 1987, (Rep. NTSB- RAR-88-01). *Washington, DC: Author.*
56. Nomura, T., Higuchi, K., Yu, H., Sasaki, S., Kimura, S., Itoh, H., Taniguchi, M., Arakwawa, T., & Kawai, K. (2006). Slow-wave photic stimulation relieves patient discomfort during esophagogastroduodenoscopy.
57. Ossebaard, H. C. (2000). Stress reduction by technology? An experimental study into the effects of brain machines on burnout and state anxiety. *Appl Psychophysiol Biofeedback*, 25(2), 93-101.
58. Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39, 230-253.
59. Patrick, G. J. (1996). Improved neuronal regulation in ADHD: An application of 15 sessions of photic-driven EEG neurotherapy. *J Neurother*, 1(4), 27-36.
60. Pfurtschellera, G., Lopes, F. H., & Silva, D. A. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*.
61. Phelps, Elizabeth A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. *Current opinion in Neurobiology*, 14(2), 198-202.
62. Rempel, J. K. & Holmes, J. G. (1986). How do I trust thee? *Psychology Today*, February, 28-34.
63. Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, 49, 95-112.
64. Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist*, 35, 1-7.
65. Sabourin, M. E., Cutcomb, S. D., Crawford, H. J., & Pribram, K. (1990). EEG correlates of hypnotic susceptibility and hypnotic trance: spectral analysis and coherence. *International Journal of Psychophysiology*, 10(2), 125-142.
66. Sanchez, J., Rogers, W. A., Fisk, A. D., & Rovira, E. (2014). Understanding reliance on automation: Effects of error type, error distribution, age and experience. *Theoretical Issues in Ergonomics Science*, 15(2), 134-160.
67. Sanei, S., & Chambers, J. A. (2007). EEG signal processing. *John Wiley & Sons Ltd.*
68. Scanzoni, J. (1979). Social exchange and behavioral interdependence. In: Burgess, R. L. & Huston, T. L., Eds. *Social Exchange in Developing Relationships*. New York: Academic Press.
69. Semrud-Clikeman, M., & Ellison, Anne. Teeter. (2009). Child Neuropsychology Assessment and Interventions for Neurodevelopmental Disorders. *New York.*

70. Seong, Y., & Bisantz, A. M. (2008). The impact of cognitive feedback on judgment performance and trust with decision aids. *International Journal of Industrial Ergonomics*, 38(7), 608-625.
71. Sheridan, T. B. (1988). Trustworthiness of command and control systems, *IFAC Man-Machine Systems*, 427-431.
72. Siever, D. (2004). The application of audio-visual entrainment for the treatment of seasonal affective disorder. *Biofeedback*, 32 (3), 32-35.
73. Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced "complacency": Development of the Complacency-Potential Rating Scale. *The International Journal of Aviation Psychology*, 3(2), 111-122.
74. Skitka, J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, 51, 991-1006.
75. Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. *International Journal of Human-Computer Studies*, 52, 701-717.
76. Slow-wave photic stimulation relieves patient discomfort during esophagogastroduodenoscopy. *J Gastroenterol Hepatol*, 21(1 Pt 1), 54-58.
77. Starter, N., & Woods, D. (1994). Pilot Interaction with Cockpit Automation II: An Experimental Study of Pilot's Model and Awareness of the FLIGHT MANAGEMENT SYSTEM. *International Journal of Aviation Psychology* 4, 1-28.
78. Uggirala, Ananth.Gramopadhyea, Anand.K., Melloya, Brain. J., & Toler, Joe.E. (2004). Measurement of trust in complex and dynamic systems using a quantitative approach. *International Journal of Industrial Ergonomics*, 34,175-186
79. Wheless, James W., Castillo, Eduardo, Maggio, Vijay, Kim, Howard, L., Breier, Joshua, I., Simos, Panagiotis G., & Papanicolaou, Andrew C. (2004). Magnetoencephalography (MEG) and Magnetic Source Imaging (MSI), *Neurologist*:[10\(3\),138-153](#).
80. Wickramasekera, I. I. E. (1977). On attempts to modify hypnotic susceptibility: Some psychophysiological procedures and promising directions. *Annals of the New York Academy of Sciences*, 296, 143-153.
81. Williams, J. H. (2001). Frequency specific effects of flicker on recognition memory. *Neuroscience*, 104(2), 283-286
82. Williams, J., Ramaswamy, D., & Oulhaj, A. (2006). 10 Hz flicker improves recognition memory in older people. *BMC Neurosci*, 7, 21.
83. Yang, Yaling., & Raine, Adrian, A. (2009). Prefrontal structural and functional brain imaging findings in antisocial, violent, and psychopathic individuals: a meta-analysis. *Psychiatry Research*. 174 (2), 81-88.
84. Zuboff, S. (1988). In the Age of the Smart Machine: The Future of Work and Power. *Basic Books*, New York.

STUDY 3–1

PRELIMINARY STUDY ON NEUROLOGICAL MEASURE OF HUMAN TRUST IN AUTONOMOUS SYSTEMS

Preliminary Study on Neurological Measure of Human Trust in Autonomous Systems

Abstract

Autonomous systems, organizations, and interpersonal interactions are rapidly becoming more complex in a contemporary society. Trust plays a crucial role in assisting the cognitive complexity and increased uncertainty in complex systems. The previous studies focus on brain regions using fMRI and efMRI, which engage with stimuli related to trust and mistrust. This research measures and analyzes brainwaves in situations involving trust and mistrust to understand the cognitive process by using an electroencephalogram (EEG). For recording brainwaves on trust and mistrust situation, this experiment uses a word elicitation study from a previous research containing words related to trust and mistrust that were ranked by survey participants. Participants' brainwaves are recorded in real time as they are watching the selected words which evoke trust and mistrust situation. The 10-20 system of electrode placement is used for recording brainwaves and power spectrum analysis is used for brainwaves analysis. The result is that specific brainwaves can be affected by a trust or mistrust situation. A goal of this research is to identify brainwaves in trust and mistrust situations and to analyze a relationship between the brainwaves and human behaviors.

1. Introduction

Autonomous systems, organizations, and interpersonal interactions are rapidly becoming more complex in contemporary society. In these complex situations, the role of trust between humans and humans, humans and machines, and in general becomes important. A stable and well-structured environment that emphasizes order and procedure does not highly depend on trust [1]. However, complex automation systems increase unstable and uncertain situations in workplaces, and eventually the increased uncertainty will increase cognitive complexity [2]. Trust plays a crucial role in assisting the cognitive complexity and increased uncertainty in sophisticated autonomous systems. Trust can be a decision aid in dynamic and uncertain situations when it is impossible to fully comprehend the intricate automation system and when it is required to have flexible behavior for unexpected situations that procedures cannot help [2]. The degree of trust can influence system performance [3], because it affects the degree of acceptance and reliance on automation [4] and operator's strategies regarding the use of automation [5]. In order to better predict performance, it is necessary to measure situations of trust and mistrust.

There are only a few studies measuring trust using brain signals. There are previous studies investigating the correlation between the degree of trust and human facial appearance using event-related functional magnetic resonance imaging (efMRI) [6], and examining interpersonal trust in online to detect different brain regions depending on conditional and unconditional trust using hyperfunctional magnetic resonance imaging (fMRI) [7]. Another study defined a relationship between brain areas and psychological processes using functional magnetic resonance imaging (fMRI) and proved a correlation of trust and mistrust with different brain areas [8]. The previous studies focused on brain regions using fMRI and efMRI, which engage with stimuli related to trust and distrust, but they did not investigate trust and mistrust with cognitive processes. This research measures and analyzes human brainwaves in situations involving trust and mistrust to understand the cognitive process by using an electroencephalogram (EEG). An EEG can monitor, record, and measure the electrical activity of the brain in real time, making it an effective tool for researching brain

functions and activities that engage with situations of trust and mistrust. For this research, the 10-20 system of electrode placement is used for recording brainwaves, and power spectrum analysis is used for brainwave analysis.

2. Method

2.1 Participants

This research recruited 4 subjects who were over 18 years old without any cognitive or psychological problems. All were either undergraduate or graduate students and participated voluntarily. Two were male and two were female. The subjects participated under the same conditions in an isolated place with guided instructions.

2.2 Framework for Experimentation

For recording human brainwaves in situations of trust and mistrust, this experiment uses a word elicitation study from a previous study containing words related to trust and mistrust that were ranked by survey participants [9,10]. This research extracts only the first through fifteenth ranked words related to trust and mistrust from the previous research. The subject is expected to think about situations of trust and mistrust evoked by the selected words. The subject is fitted with a cap of electrodes and seated in front of a computer screen.

Before starting the experimentation, the instruction on the screen is presented for 10 seconds. Each word is presented for 7 seconds, which is enough time to read the word and understand the meaning of the word. Each slideshow is presented for 3 minutes and 30 seconds, which is enough time to evoke a situation of trust or mistrust. An experiment comprises of two sessions: 1) trust words and 2) mistrust words.

When the subject watches the first slideshow, which consists of the 30 random stimuli of the selected 15 words related to trust, the brainwaves are recorded. The subject takes a break for 1 minute to avoid fatigue and stress. When he or she watches the second word slideshow, which consists of the 30 random stimuli of the selected 15 words related to mistrust, the brainwaves are recorded.

2.3 EEG Recording and analysis

The EEG data are recorded using the g.HIamp (256 multichannel amplifier), g. GAMMAsys (active electrode system with g. GAMMAcap) and g. Recorder (brain signal recording software) by g. tec medical engineering company. According to the international 10-20 system of electrode placement, the 20 electrodes (Fp1, Fp2, Fpz, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, T6, O1, O2) are recorded and measured (see figure1). The 10-20 system is based on the relationship between the location of an electrode and the underlying area of the cerebral cortex. The letters F, T, C, P, and O represent the frontal, temporal, central, parietal, and occipital lobes. Even numbers refer to the right hemisphere, and odd numbers refer to the left hemisphere. The subject's head is fitted with a cap of electrodes (g. GAMMAcap) and the selected 20 electrodes are filled with abrasive electrolyte gel using a syringe. For the setting to record the brainwaves, the sampling frequency is set as 256 Hz, high frequency filter as 50 Hz and low frequency filter as -50 Hz.

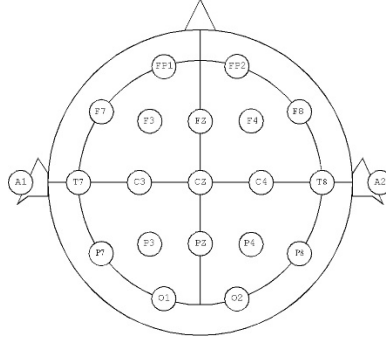


Figure 1: International 10-20 system for electrode placement

3. EEG Analysis

For analyzing human brainwaves, there are linear spectral measurements, nonlinear measurements of complexity, and interdependency measurements [11]. As one of the linear spectral measurements, the power spectrum measurement is a common method to quantify EEG data to present the distribution of signal power by frequency [12].

To identify specific brainwaves in situations of trust and mistrust, the recorded raw EEG data are analyzed by the power spectrum through the EMSE Suite Data Editor by Cortech Solutions. In the power spectrum analysis, an FFT (Fast Fourier Transform) is performed on defined data intervals and the results are squared and then averaged. The units are Volts². The entire time series file from the whole experimental time is selected and analyzed by the power spectrum. The power spectrum data is displayed in units of Volts²(V²). The axis and legend scaling is shown in units of uV²(uV²). EMSE uses the mathematical convention in which uV² is an equivalent representation of (uV)². See equation (1) below.

$$1.0 \text{ uV}^2 = 1.0 (\text{uV})^2 = 1.0e^{-0.12\text{V}^2} \quad (1)$$

4. Result

In this research, the subjects' brainwaves are recorded as they respond to the selected word related to trust and mistrust by the 10-20 system of electrode placement (22 channel) with a total time of 210 seconds for each session. They are then analyzed by power spectrum analysis. The brainwaves of the 4 subjects are normal, without any extreme brain activity, so the average of all collected brainwaves is used for analysis.

This research focuses on identifying specific brainwaves and cognitive processes related to situations of trust and mistrust. Using power spectrum analysis, the recorded brainwaves are categorized by their frequencies: alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-60 Hz). The beta waves are divided into low beta (13-15 Hz, which is called as SMR), mid beta (15-20 Hz) and high beta (20-30 Hz) for more detailed analysis [13,14].

Using the power spectrum, the power of alpha waves is averaged from 8 Hz to 13 Hz. In alpha waves, the average of trust is 4.23 uV² and the average of mistrust is 1.90 uV², so the difference is 2.33 uV² (see Figure 2). The power of beta is averaged from 13 Hz to 30 Hz. The average of trust is 4.28 uV² and the average of mistrust is 2.61uV², so the difference is 1.67 uV² (see Figure 3). The power of gamma is averaged from 30 Hz to 60 Hz. The average of trust is 2.89 uV² and the average of mistrust is 3.19 uV², so the difference is -0.3 uV²,

indicating that mistrust is higher than trust (see Figure 4). Therefore, alpha and beta waves are stronger in trust situations and gamma waves are stronger in situations of mistrust.

Situations of trust and mistrust are compared by low, mid, and high beta waves, which are related to cognitive processes. The power of low beta is averaged from 13 Hz to 15 Hz. The average of trust is $4.62 \mu V^2$ and the average of mistrust is $2.10 \mu V^2$, so the difference is $2.52 \mu V^2$. The power of mid beta is averaged from 15 Hz to 20 Hz. The average of trust is $4.39 \mu V^2$ and the average of mistrust is $2.54 \mu V^2$, so the difference is $1.85 \mu V^2$. The power of high beta is averaged from 20 Hz to 30 Hz. The average of trust is $2.82 \mu V^2$ and the average of mistrust is $2.88 \mu V^2$, so the difference is $-0.06 \mu V^2$, indicating that the mistrust is higher than the trust (see Figure 5). Therefore, low and mid beta waves are stronger in situations of trust and high beta waves are stronger in situations of mistrust. The difference between the averages of trust and mistrust decreases from low beta ($2.52 \mu V^2$) to high beta ($-0.06 \mu V^2$).

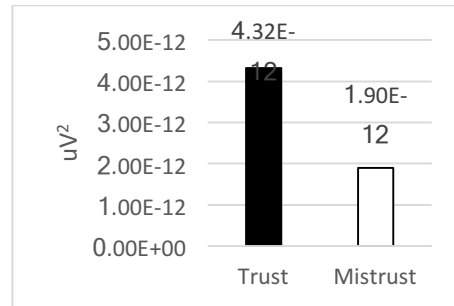


Figure 2: Comparison trust and mistrust in α waves

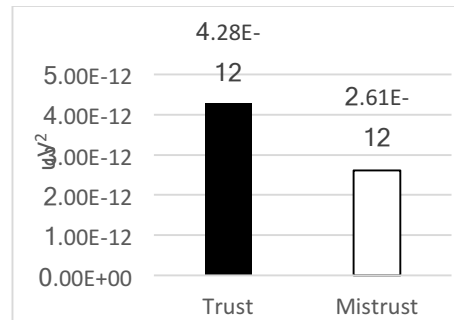


Figure 3: Comparison trust and mistrust in β waves

5. Conclusion

Based on the power spectrum analysis, situations of trust and mistrust are compared by α , β , and γ waves. While alpha and beta waves are stronger in situations of trust, gamma waves are stronger in situations of mistrust (see Figure 2-4). Alpha waves are related to a calm state without any stress or tension [15], and gamma waves are related to complicated cognitive processes with significant stress and anxiety, so trust situations can help to calm down any mental activity, but situations of mistrust can interrupt mental activity because of increased stress and anxiety.

This research focuses on the beta waves because they are detected in the normal waking state of consciousness related to cognitive tasks [16]. Beta waves can affect cognitive tasks such as attention, problem solving, judgment, and decision-making [17]. In sustained attention tasks, SMR affects perceptual sensitivity and lowers omission errors and mid beta waves (beta 1) reduces reaction times [16,18]. Beta waves can be a guideline to measure cognitive process [19]. High beta waves are associated with intensity, anxiety and hypervigilance [17].

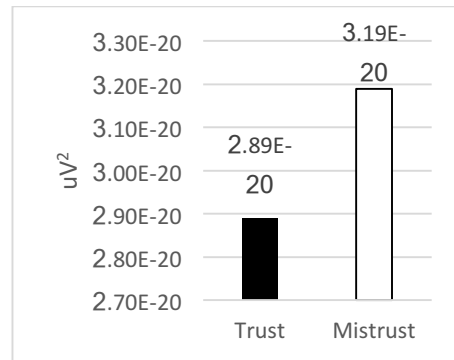


Figure 4: Comparison trust and mistrust in γ waves

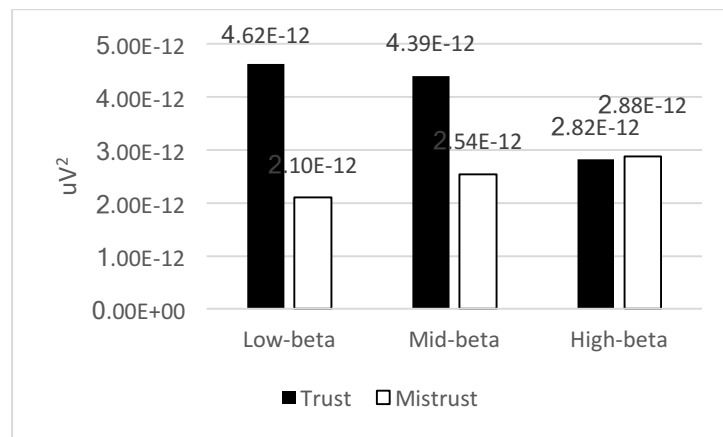


Figure 5: Detailed comparisons of trust and mistrust in β waves

Overall, all beta waves are more powerful in situations of trust than situations of mistrust. On a closer view, in trust situations, the power of beta waves decreases from low beta to high beta. In contrast, in situations of mistrust, the power of beta waves increases from low beta to high beta. (see Figure 5). Therefore, trust situations can aid sustained attention and problem solving, which are related to low beta and mid beta waves, but situations of mistrust can disturb the focused mental activities exposed by intensity and anxiety, which are related to high beta waves.

6. Discussion

According to this research, trust can assist in effective decision-making by increasing concentration and performance, but mistrust can interrupt effective decision-making by increasing stress and anxiety. Even though there were limited samples of brainwaves, this

research presents the difference between trust and mistrust. Further research with more samples of brainwaves will aid in the further analysis of the differences between trust and low and mid beta waves and mistrust and high beta waves to define how trust can affect performance and decision-making.

References

1. Moorman, Christine., Rohit, Deshpande., and Gerald, Zaltman., 1993, "Factors Affecting Trust in Market Research Relationships," *Journal of Marketing*, 57(January), 81-101.
2. Lee, J.D., See, Katrina. A., 2004, "Trust in Automation: Designing for appropriate Reliance," *Human Factors and Ergonomic Society*, 46(1), 50-80.
3. Ananth Uggirala, Anand K., Gramopadhyea, Brain J. Melloya., and Joe, E. Tolerb., 2004, "Measurement of trust in complex and dynamic systems using a quantitative approach," *Inter. Journal of Industrial Ergonomics* 34(3), 175-186.
4. Sheridan, T.B., 1988, "Trustworthiness of command and control systems," *IFAC Man-Machine Systems*, 427-431.
5. Lee, J.D., Moray, N., 1994, "Trust, control strategies and allocation of function in human-machine system," *Ergonomics*, 35, 1243-1270.
6. Winston, J.S., Stranger, B.A., O'Doherty, J., and Dolan, R.J., 2002, "Automatic and intentional brain responses during evaluation of trustworthiness of faces," *Nature Neuroscience*, 5(3), 277-283.
7. Krueger, Frank., McCabe, Kevin., Moll, Jorge., Kriegeskorte, Nikolaus., Zahn, Roland., Strenziok, Maren., Heinecke, Armin., and Grafman, Jordan., 2007, "Neural correlates of trust," *PNAS*, 104(50), 20084-20089.
8. Dimoka, Angelika., 2011, "Brain mapping of psychological processes with psychometric scales: An fMRI method for social neuroscience," *NeuroImage*, 54(1), S263-S271.
9. Jian, Jiun-Yin., Bisantz, Ann. M., Drury, Colin. G. 2000, "Foundations for an Empirically Determined Scale of Trust in Automated Systems," *International journal of cognitive ergonomics*, 4(1), 53-71.
10. Bisantz, Ann. M., Finger, Richard., Seong, Younho., and Llians, James., 1999, "Human Performance and Data Fusion Based Decision Aids," *International Society of Information Fusion*.
11. Teplan, M., Krakovska, A., Stolc, S., 2006, "EEG responses to long-term audio-visual stimulation," *International Journal of Psychophysiology*, 59(2), 81-90.
12. Dressler, O., Schneider, G., Stockmanns, G., and Kochs, E.F., 2004, "Awareness and the EEG power spectrum: analysis of frequencies," *British of Journal Anaesthesia*, 93(6), 806-9.
13. Sanei, S., and Chambers, J.A., 2007, "EEG signal processing," John Wiley & Sons Ltd.
14. Pfurtschella, G., Lopes, F.H., Silva, Da., 1999, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*.
15. Ashtaputre-Sisode, Aparna., 2016, "Emotions and Brainwaves," *The international Journal of Indian Psychology*, 3(2), No5.
16. Egner, T., and Gruzelier, J. H., 2004, "EEG biofeedback of low beta band components: Frequency-specific effects on variables of attention and event-related brain potentials," *Clinical Neurophysiology*, 115(1), 131-139.
17. Patrick N. Friel, BS., 2007, "EEG Biofeedback in the Treatment of Attention Deficit/ Hyperactivity Disorder," *Alternative Medicine Revie*, 12(2).
18. Lqbai, Sadaf., Muhammed, Shanir P.P., Khan, Yusuf. Uzzaman., and Farooq, Omar., 2016 "EEG Analysis of Imagined Speech," *International Journal of Rough Sets and Data Analysis*, 3(2).
19. Ray, W.J., and Cole, H.W., 1985, "EEG Alpha Activity Reflects Attentional Demands, and Beta Activity Reflects Emotional and Cognitive Processes," *Science*, 228 (4700), 750.