THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

COLLEGE OF INFORMATION SCIENCES AND TECHNOLOGY

# PENN STATE RED CELL ANALYTICS LAB GRAND CHALLENGE: A STUDY ON ANALYTIC BIAS

MELISSA MARIE SIRERA
SPRING 2014

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Security and Risk Analysis and Global and International Studies
with honors in Security and Risk Analysis

Reviewed and approved* by the following:

Jacob Graham
Colonel, USMC (Ret)
Professor of Practice in Security and Risk Analysis
Thesis Supervisor

Andrea Tapia
Associate Professor of Information Sciences and Technology
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

## ABSTRACT

This thesis examines the totality of the process of designing and conducting an Institutional Review Board (IRB) approved human-subjects experiment with the purpose of identifying judgment bias and measuring the effects of structured techniques on mitigating analytic bias. The study utilized an analytic decision game developed by Colonel Jacob Graham, USMC (ret) to test for bias in intelligence analysis. The research, conducted in partnership with Raytheon Company, adapted the experimental design utilized in the Raytheon Grand Challenge. The study evaluated the effectiveness of three structured analytic techniques: Link Analysis, Analysis of Competing Hypotheses (ACH), and Information Extraction and Weighting (IEW). Researchers compared these techniques to a control group to determine the impact of the structured analytic techniques on mitigating bias that may occur during decision-making. After establishing if this relationship existed, researchers sought to determine which structured analytic tool was most effective in reducing judgment bias. The study will complement the literature available on analytic bias and explore areas of bias reduction that current literature has not yet investigated. Conclusions drawn from this study can help analysts understand the effects of judgment bias during analytic decision-making and the impact of structured analytic techniques on reducing bias. The overall goal of this research is to contribute literature that will help improve the analytic products and judgments that influence decision making.

# TABLE OF CONTENTS

## LIST OF FIGURES

# LIST OF TABLES

## ACKNOWLEDGEMENTS

I would like to thank Colonel Jacob Graham, USMC (ret) for his mentorship the past three years. Colonel Graham has positively impacted my education as a teacher, thesis supervisor, and club advisor. As my thesis supervisor, Colonel Graham invited me to work on a research project that would help me to produce a meaningful honors thesis and prepare me to succeed in my future career endeavors. Colonel Graham has always been available as a resource, to address any questions or concerns I had, throughout the thesis research and writing process. Additionally, Colonel Graham worked to ensure that I had the equipment and materials necessary to conduct the research experiments. Without the help of Colonel Graham, successful completion of this thesis would not have been possible. I cannot thank him enough.

Secondly, I would like to thank Donald Kretz, B.J. Simpson, and Raytheon Company for allowing me to adapt their experiment design. The experiment framework and materials were an invaluable resource. Furthermore, the expertise and advice of Mr. Kretz were a tremendous help. Without Raytheon Company, the research portion of this thesis would not have been possible.

I would also like to thank the Red Cell Analytics Lab executive board for assisting me during the facilitation of the analytic exercises. Specifically, I would like to thank Emily Catherman for attending all of the experiment sessions, helping to ensure that everything ran smoothly, and assisting me with anything I needed. Emily was an enormous help.

Additionally, I would like to thank the College of Information Sciences and Technology for supporting me in this research. I would like to thank Dr. Nick Giacobe for his advice and help during the IRB process and Dr. Katherine Hamilton for sharing her expertise in research methods and statistical analysis. Furthermore, I want to thank the College of IST HelpDesk for providing me with the necessary software and equipment for this study. My final thank you is to all the subjects who volunteered to participate in my study.

# Chapter 1

# Introduction

According to the United States Department of Defense, a total of 4,489 U.S. military and civilian personnel were killed and 32,237 military personnel were wounded during Operation Iraqi Freedom and Operation New Dawn. Operation Iraqi Freedom, a campaign of U.S. combat operations in Iraq, began on March 19, 2003 and ended on August 31, 2010 (U.S. Department of Defense, 2014). Operation New Dawn commenced immediately after the end of Operation Iraqi Freedom; this new mission shifted the U.S. troop focus from combat operations to advising and supporting Iraqi security forces. On December 15$^{th}$, 2011, Operation New Dawn ended, signifying the official conclusion of the U.S. War in Iraq (Torreon, 2012).

In 2011, a scholarly research group from the Watson Institute for International Studies at Brown University released an extensive report on the financial, human, and civil impacts of the War in Iraq. More than 30 individuals--including anthropologists, economists, humanitarian personnel, lawyers, and political scientists--have contributed to this research initiative, the Costs of War Project. In March of 2013, the research group released updated research findings in relation to the 10$^{th}$ anniversary of the U.S. Invasion of Iraq. Their study found that between 2003 and 2013, more than 189,000 direct deaths resulted from the Iraq War. These deaths included: civilians, opposition forces, allied military and police, other allied forces, U.S. contractors, U.S. military, humanitarian workers, and journalists (Costs of War Project, 2013). Figure 1 illustrates the number of deaths for each demographic and the proportion of deaths to the total death toll. Neta C. Crawford, a professor of political science at Boston University, estimates that the number of Iraqi civilians that have been severely wounded due to the Iraq War is close to the number of civilian casualties (Crawford, 2013). Statistics on both casualties and injuries can offer a quantitative description of the consequences of war; however, the impacts felt as a result of the loss of human lives can never be fully defined.

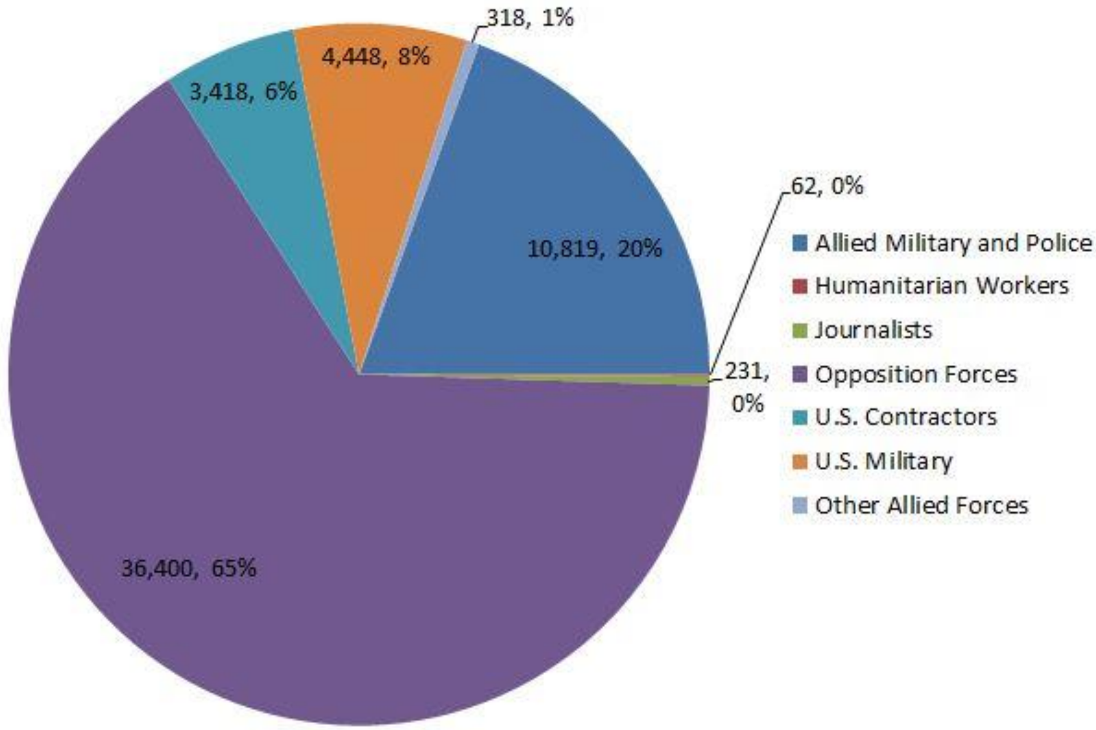


**Figure 1 Pie chart of deaths directly related to the Iraq War. Adapted from "Iraq: 10 Years After Invasion" by Costs of War Project, Watson Institute for International Studies, Brown University, 2013.**

The effects of war, although not as emotionally significant, extend beyond the human element. The United States has seen, and will continue to see in the future, large-scale economic impacts as a result of the Iraq War. According to a March 2011 Congressional Research Services report, an estimated total of $806 billion dollars had been allocated for the war in Iraq. This estimate was based on the 6th Continuing Resolution for the 2011 fiscal year, as well as congressional budget submissions from Department of Defense, Department of Veterans Affairs, and Department of State and U.S. Agency for International Development (Belasco, 2011). In its March 2013 report, the aforementioned research initiative, The Costs of War Project, stated that the United States spent approximately $1.69 trillion through fiscal year 2013. The Costs of War Project calculated the total U.S. Iraq war expenses from financial information reported in federal budgets. The calculation included budgetary information from the Department of Defense, Department of State and the U.S. Agency for International Development, Department of Veterans Affairs, Social Security Administration, and Department of Homeland Security. The cost estimate also included the cost of interest on the money borrowed to fund the war (Crawford, 2013). Table 1 shows a breakdown and summary of the total costs of the Iraq War, through 2013, for the United States.

**Table 1 U.S. Total Costs of Iraq War. Adapted from *U.S. Costs of Wars Through 2013: $3.1 Trillion and Counting: Summary of Costs for the U.S. Wars in Iraq, Afghanistan and Pakistan* (p. 11), by Neta C. Crawford, 2013, Watson Institute for International Studies, Brown University.**

| U.S. Costs of Iraq Through FY2013 | USD (Billions) |
|---|---:|
| DOD War Appropriations | 769.9 |
| State/USAID | 52.38 |
| Increase in DOD Base Spending (attributing 54% of total increase to Iraq | 401.27 |
| VA Medical (attributing 65% of total costs to Iraq War) | 15.31 |
| VA Medical (attributing 65% of total cots to Iraq war) | 21.84 |
| SS Disability (attributing 65% of total costs to Iraq war) | 2.91 |
| Growth in other VA Spending (attributing 65% of total costs to Iraq) | 46.48 |
| Homeland Security (attributing 54% of increase to Iraq) | 245.81 |
| **Total Cost of Iraq War and War-Related Expenses through 2013** | **1,694.70** |

Beyond the $1.69 trillion dollars already spent on the U.S. War in Iraq, the Cost of War Project also estimated that the United States will be required to pay approximately $490 billion for Veterans care and disability between 2014 and 2053. The United States, by 2053, will pay at least an estimated $2.18 trillion in total as a result of the direct and indirect expenses of the Iraq

War. This estimate does not take into consideration the amount of interest the United States will need to pay until 2053 (Crawford, 2013). The financial and demographic statistics of the U.S. War in Iraq demonstrate the high prices our nation paid as a result of one of the largest United States Intelligence Community failures.

Prior to the war in Iraq, the United States Intelligence Community (IC) assessed that Saddam Hussein had rebuilt, and was continuing to grow, Iraq's nuclear weapons program. Specifically, the analysts in the U.S. IC assessed that Hussein possessed biological weapons, had the ability to produce biological weapons in transportable facilities, stored a surplus of chemical weapons, and was currently manufacturing more weapons of chemical warfare (The Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction, 2005). Former President George W. Bush commissioned a committee, in 2005, to conduct a thorough review of the assessments and capabilities of the United States Intelligence Community. In the letter to the president that prefaced their report, the Commission wrote, "We conclude that the Intelligence Community was dead wrong in almost all of its pre-war judgments about Iraq's weapons of mass destruction" (The Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction, 2005). This conclusion expressly reveals the extent of the Intelligence Community's inaccuracy. This type of failure was atypical for the United States; however, the damage caused by the failure and to the United States was severe. An excerpt from the *Report to the President of the United States* reads:

> While the Intelligence services of many other nations also thought that Iraq had weapons of mass destruction, in the end it was the United States that put its credibility on the line, making this one of the most public—and most damaging—intelligence failures in recent American history (The Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction, 2005).

As the quote describes, the United States Intelligence Community's wrongful assessment of Iraq's nuclear weapons program is one of the best examples, in recent history, of a failure in intelligence analysis. A failure, such as this, reminds the entire world of the importance of accurate intelligence analysis.

Beyond clarifying the degree of inaccuracy, the Commission also explains the reasons why the failure occurred – the analysts exhibited bias during the analytic process. The report states, “This failure was in large part the result of analytical shortcomings; intelligence analysts were too wedded to their assumptions about Saddam's intentions” (The Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction, 2005). As explained in the government report, analysts clung to their pre-conceived notions that Saddam Hussein intended to develop and use weapons of mass destruction. This intelligence failure demonstrates the existence and harmful effects of confirmation bias and expectation bias in the field of intelligence analysis.

One of the most troubling observations of the Commission is the fact that analysts were unable to identify their bias. The Commission wrote to the president, in the letter that accompanied the report, “What the intelligence professionals told you about Saddam Hussein’s programs was what they believed. They were simply wrong” (The Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction, 2005). The inability to properly identify bias allows for the future possibility of flawed intelligence assessments. For this reason, researchers must work to not only develop techniques to mitigate bias, but they must also establish methods to identify bias.

The experiments conducted by both the Raytheon Grand Challenge and the Red Cell Analytics Lab (RCAL) Grand Challenge help contribute to the field of research on analytic bias. It is essential that bias is reduced in analysis. The work of analysis involves the process of examining the multiple, complex pieces of a problem in an effort to find the best solution; this process is inherently difficult. As the quantity of problems the U.S. Intelligence Community needs to solve increases, their task, as a whole, becomes more difficult. The consequences of flawed intelligence assessments are extremely severe; thus, it is highly important that researchers find ways to both identify bias and develop techniques to successfully mitigate the effects of bias on analytic products when they occur.

# Chapter 2

## Literature Review

The intelligence analysts in the United States Intelligence Community showed signs of confirmation bias in their assessment on Iraq's possession of and intent to use weapons of mass destruction. Confirmation bias is widely discussed within psychological literature, and many researchers have conducted empirical studies in order to better understand the ways to reduce this cognitive problem. Lehner, Adelman, Cheikes, and Brown (2008) conducted a study to investigate the effectiveness of Analysis of Competing Hypotheses (ACH) in reducing confirmation bias in a complex analysis task similar to intelligence analysis, financial analysis, and law enforcement investigations. Researchers examined the impact of participants' use of the Analysis of Competing Hypotheses procedure and participants' professional intelligence analytic experience on reducing confirmation bias.

Confirmation bias was measured as a resistance to the trends supported by the study's presented evidence. Researchers used a sample of employees of a Research and Development company and randomly assigned participants into two experimental conditions: ACH and Non-ACH. The participants' objective was to determine the source of an explosion on the USS Iowa from three hypotheses. Participants received five stages of evidence that either confirmed or disconfirmed one of the three hypotheses. Throughout the experiment, participants rated the extent that each piece of evidence supported, or did not support, each hypothesis and their belief in each hypothesis. The researchers measured confirmation bias in two ways: the extent to which participants did not change their beliefs when presented with evidence that contradicted their previous beliefs and the extent that a prior belief in a hypothesis impacted the assessment of new evidence. The results showed that there was the presence of confirmation bias in the experiment and that the use of the ACH method by participants without prior intelligence analysis experience greatly reduced confirmation bias exhibited during evidence assessment. However, there was no evidence that the use of ACH reduced confirmation bias exhibited by experienced analysts. The researchers concluded that future research should continue to investigate the use of ACH in decreasing confirmation bias in analytic tasks.

This research has several limitations. First, the researchers used a convenience sample. Researchers relied on those employees that volunteered as opposed to ensuring that the sample was representative of the entire company. Second, the researchers did not ensure that each experimental condition had an equal amount of non-experienced and experienced participants. This inequality may have affected the results because researchers had a disproportionate amount of data for non-experienced participants in the ACH condition and experienced participants in the ACH condition. The results showed that confirmation bias was reduced with the use of ACH by participants who lacked experience; however, there was more data collected for this group because there was a larger amount of participants without experience in the ACH condition. Within the article, the researchers discussed that the participants may have developed different goals based on the differences in task completion of the two experiment conditions. For example, the non-ACH group may have been working to develop a coherent story; whereas, the ACH group may have been more focused on evaluating each piece of evidence in relation to a specific hypothesis. This aspect of the experimental design may have impacted the results and should be investigated by future researchers. Lehner and colleagues (2008) represent a good starting point from which future research can build by continuing to develop experiments that study the effectiveness of ACH and other structured techniques in reducing confirmation bias for experienced intelligence analysts.

Cook and Smallman (2008) conducted an empirical study to examine the effectiveness of two debiasing techniques in reducing confirmation bias. This experiment evaluated a new collaborative analytic tool called Joint Intelligence Graphical Situation Awareness Web (JIGSAW). The study compared the effectiveness of the graphical format of the JIGSAW visual evidence layout versus information displayed in a text format in reducing confirmation bias. The study also examined the effect of the foundation of evidence assessment on decreasing confirmation bias by comparing a decision-maker who had access to other analysts' evidence assessments and a decision-maker only working with their individual evidence assessments. Participants were assigned to experimental conditions then tasked to evaluate, choose, and rank evidence from an evidence pool that contained an equal amount of supporting and disagreeing evidence.

The dependent variable, confirmation bias, was measured in three ways; each phase had its own measurement. During the evaluation phase, confirmation bias was measured by the average amount of supportive ratings assigned to evidence by the participant. The selection phase showed confirmation bias if the participant did two things: 1) if the average of participant selected evidence had a supportive rating and 2) if the participant generally selected a larger amount of supportive evidence. The final phase showed an existence of confirmation bias if the participants assigned a more important ranking to evidence with greater evaluation scores.

For the first phase, the results showed that in comparison to the judges' assessments, the participants were unbiased in evaluating the evidence, and the participants' assessments were not influenced by the format of the evidence layout. Both the "selection" and "ranking" sections, found that all conditions exhibited confirmation bias; however, the graphical evidence format significantly reduced the bias. The evidence source had no effect on the amount of confirmation bias.

The first potential limitation occurs with the representativeness of the sample. Twenty percent of this sample was comprised of women. The results found in this study may therefore only be applicable to males. This may represent a threat to the study's external validity because the results might not be generalizable to females in the target population. Another potential threat to external validity may have occurred with the sample being restricted to Naval analysts. The results may be different if more branches of the military, as well as various agencies within the U.S. Intelligence Community, had been included in the sample. Based on their place of employment, analysts may receive different amounts and styles of training; therefore, graphical evidence layouts and the evidence source may be more or less effective in reducing bias as a result of the analyst's employer.

The article also discussed the limitation of having the "other analysts' assessments" being presented through graphic and text formats, as opposed to being presented in person. The assessments also lacked information about the other analysts' experience and credibility. If multiple analysts had been collaborating together, or if more information had been provided about the "other analysts", then the variable of "other analysts' evidence assessments" may have had a greater effect on reducing confirmation bias. Researchers acknowledged that it would be very beneficial to conduct this experiment again in a more collaborative environment. I also think it would be important for researchers to perform another experiment that provides information about the other analysts' credibility. Another proposed area of future research, as suggested in the study, was to incorporate the evaluation of multiple hypotheses within the graphical evidence layout.

Lehner, Adelman, DiStasio, Erie, Mittel, Olson (2009) conducted an empirical study on confirmation bias in technical data analysis and on a method to reduce this bias through consideration of alternative causes. Researchers wanted to focus on confirmation bias within analysis of remote sensing data because there are very few empirical studies on confirmation bias in technical data analysis. Furthermore, technical analysis of remote sensing data is very important in helping to solve many military, intelligence, and emergency relief problems.

The study had two independent variables: types of experimental problems (supportive or non-supportive of the hypothesis) and the use of a sensor analysis template, which required analysts to consider alternative causes. Participants were randomly assigned into two groups; one group used the sensor analysis template and the other did not. Researchers asked participants to individually complete six analytic problems; half of the questions had correct answers that supported the hypothesis, while the other half disconfirmed the hypothesis. Confirmation bias was measured as a pattern of analysts falsely identifying sensor data as confirming a hypothesis. The researchers determined that their alternative cause approach helped to reduce bias. The combination of results from the study allowed researchers to conclude, in general, that technical data analysts can decrease sensor interpretation errors by using methods that encourage analysts to spend time identifying and evaluating alternative causes.

This study had a potential limitation in the sample. A small convenience sample of only ten participants was used in the study; this sample size is a threat to external validity in that a small sample makes it difficult to generalize the results of the study to the larger target population of technical analysts. However, the primary potential limitation with this study is that the participants worked individually. According to the study, most technical analysts have their analytic decisions reviewed by their supervisors and peers. This aspect of technical analysis suggests that future research should investigate how the use of the alternative causes method impacts confirmation bias in accordance with an organization's review process.

Jonas, Schulz-Hardt, Frey, and Thelen (2001) conducted four empirical studies on confirmation bias in seeking information after a decision has been made. These studies were conducted because there are very few empirical studies that have produced evidence of confirmation bias occurring during the information search process. I will review the first experiment, which involved researchers seeking to determine if confirmation bias exists in both sequential and simultaneous information seeking. Furthermore, researchers wanted to investigate if the two processes differed in terms of the strength of the bias. The two experimental conditions for the independent variable of type of search mode were simultaneous information seeking and sequential information seeking. The two types of information collected were either supporting or conflicting.

In this study, confirmation bias is defined as the participant's preference of choosing supporting information over conflicting information. This bias is measured by the difference between the number of disconfirming articles that were selected and the number of supportive articles that were selected. Participants were randomly assigned to an experimental condition; each participant had to complete a three part questionnaire involving initial decision-making, information selection, and final decision-making. The study's results revealed that confirmation bias was exhibited by participants using both simultaneous and sequential information search methods. However, the results also showed that the amount of confirmation bias that occurred in sequential information searching was significantly more than the amount of bias that occurred in the simultaneous information seeking process.

The limitations of this study can help serve as a guide for future research. Participants were asked to make a decision regarding health insurance coverage; however, the authors explained that there was an increased public awareness about health insurance policies prior to conducting this experiment. Building on this study, future experiments should further investigate if prior knowledge, or a lack of prior knowledge, impacts the amount of confirmation bias exhibited by participants. The researchers also report that participants' age and gender did not interfere in the results. Future research in this area should attempt to replicate the findings among various demographic groups.

Wheeler and Arunachalam (2008) conducted an empirical study on confirmation bias exhibited by tax professionals during tax research and the effects of decision aid methods on this bias. There are few empirical studies that investigate the effects of specific designs of decision aids for decision-making involving taxes. For this reason, the purpose of the research was to determine which type of decision design aids is most effective in reducing confirmation bias demonstrated by tax professionals during tax research. Participants were randomly assigned to three experimental conditions: decision aid with a design requiring justification, decision aid with an evaluation checklist design, and no decision aid. The task involved participants completing a seven part questionnaire that mimicked a tax research project concerning whether a bonus could be recorded as a deduction for a corporation's taxes. Early in the task, participants were asked to tentatively decide a solution, prior to conducting research.

Confirmation bias was measured by the number of tax research cases selected by the participant that supported their initial judgment and the participants' evaluations of the importance of the cases that supported their position. The results showed that the design involving justification significantly reduced confirmation bias in comparison to the unaided and evaluation checklist conditions. The results for the evaluation checklist design show that this design either increased confirmation bias or had no effect on the amount of confirmation bias in comparison to the unaided method and the method using a justification design decision aid. Furthermore, researchers concluded that these design aids, although tested in paper format, would work in digital formats and could be applied to all methods of tax research.

The first limitation of this study involved the design of the decision aids. Although researchers designed each method based on the common characteristics of these decision aids, researchers could have developed the designs of each tax decision aid differently. For example, the evaluation checklist had a format that focused on an individual evaluation of predetermined factors; however, researchers could have developed a design that guided the participants through the process of decision making while also encouraging them to evaluate the factors in relation to one another. Researchers also could have changed the design of the judgment decision aid in a way that would cause participants to not carefully evaluate the justifications behind their decisions. These limitations suggest that the results of this study might not be generalizable to all decision aid evaluation checklist designs or justification requirement designs. The researchers suggest that future experiments could involve improving the designs of both decision aids to reduce the weaknesses and increase the strengths of each aid.

The study was also limited in that it only measured confirmation bias after the client had selected a tax position. The experiment task involved researchers in the planning stage of tax decision making; however, they did not measure the dependent variable in this stage. Future research could study how decision aids affect the presence of confirmation bias in the planning stage of tax research.

Another potential limitation is that the study only examined paper-based decision designs; however, the researchers still believed principles of the study's designs could be implemented in a digital format. Future research may benefit from testing these principles when applied in an electronic decision aid. The article mentioned that there are already a large number of digital decision aid technologies that are designed to help improve tax research. Future research should examine the effectiveness of these digital technologies in reducing confirmation bias in tax research.

# Chapter 3

## Raytheon Grand Challenge

A collaborative research initiative between Raytheon Company and the Pennsylvania State University College of Information Sciences and Technology (IST) provided the framework for the Penn State Red Cell Analytics Lab (RCAL) Grand Challenge. The combined research effort, which began in 2011, focused on user-centric analytics. User-centric analytics are defined as analytics that identify and describe the activity, environments, and behavioral patterns of analysts. Raytheon and the College of IST collaborated on the user-centric analytics research in an effort to boost the analytic intelligence yield from an intelligence analyst (Hall & Graham, 2011).

In order to improve analytic yield, Raytheon conducted experiments to study factors that influence analytic performance and the effectiveness of analytic techniques in improving the quality of analysis. The College of IST provided Raytheon with the Analytic Decision Game (ADG) methodology to support their experiment. Colonel Jacob Graham, Professor of Practice in the College of IST, developed the ADG methodology as a method to teach decision-making theory, collaboration, and analysis. This methodology presents challenging problems in a simulated scenario-based analytic gaming environment (Hall & Graham, 2011).

Analytic decision game simulations are comprised of synthetic data elements; these data elements include observations, intelligence products, and reports. A probable course of action or logical story comes together when an individual considers these data elements as a whole from end-to-end (Kretz, Simpson, & Graham, 2012). The Analytic Decision Game methodology helps researchers investigate gaps in understanding and cognition that occur during the analytic process. Furthermore, the insights realized by the use of the ADG approach can be used to develop cognitive tools to strengthen situational awareness and analytic quality (Hall & Graham, 2011). Raytheon used an analytic decision game, created by the College of IST, to conduct an analytic exercise that simulated the analysis process of United States military intelligence analysts (Kretz, Simpson, & Graham, 2012).

The scenario for the Raytheon Grand Challenge takes place in the year 2010 and asks participants to conduct intelligence analysis on Improvised Explosive Device (IED) attacks that occurred near the Green Zone in Baghdad, Iraq. Although the exercise involves a fictitious scenario, the data set represents a realistic situation. The data elements for this exercise were provided by the College of IST from their Synthetic Counterinsurgency (SYNCOIN) data set. Graham developed the SYNCOIN data set as the Pennsylvania State University contribution to the Army Research Office sponsored Multidisciplinary University Research Initiative (MURI). This multi-year research project created the SYNCOIN data to aid the development of tools and techniques developed to enhance Situation Awareness and Situation Understanding. These methods included: analysis and data fusion processes, interfaces and process flows for user-in-the loop analysis and automated process analysis, and data stores design. The synthetic data elements utilized by Raytheon for its Grand Challenge experiment were only a small portion of the entire SYNCOIN data set (Kretz, Simpson, & Graham, 2012).

With its Grand Challenge experiment, Raytheon attempted to identify various types of judgment biases that may occur during the analytic process. Behavioral scientists observed that bias occurs in forensic analysis when analysts examine an evidence collection. Analysts influenced by judgment bias have a tendency to choose weaker conclusions as opposed to conclusions more thoroughly supported by the evidence. There are multiple types of bias that can lead to this judgment error; these biases are confirmation bias, expectation bias, and anchoring effect. Confirmation bias occurs when a person places a higher weight on information that confirms his or her existing beliefs. Similarly, expectation bias is the tendency of an individual to place higher consideration on the evidence that agrees with his or her expectations for a specific outcome. Anchoring effect refers to the disposition to place greater emphasis on a specific piece of evidence or past reference. The Raytheon Grand Challenge experiment sought to investigate the judgment biases of confirmation bias and anchoring effect. Raytheon researchers also sought to measure the effectiveness of analytic multipliers in reducing these decision biases. The analytic multipliers included the techniques of Link Analysis, Analysis of Competing Hypotheses (ACH), and Information Extraction and Weighting (IEW). Raytheon used three experimental conditions to evaluate these techniques (Kretz, Simpson, & Graham, 2012).

The Grand Challenge Experiment was comprised of three segments: pre-game, game phases, and post-game. During the pre-game segment, participants filled out a background and skills questionnaire. Raytheon used information collected in the background questionnaire to arrange participants into groups, based on analytic experience and subject matter knowledge, to utilize the randomized block technique for participant assignment to the various experimental conditions. After the participants were organized into blocks, researchers performed random sampling to assign each participant to one of the experimental conditions. Due to the conflicts with the Raytheon work schedule, not all subjects were available to participate on their assigned day and researchers had to allow the assignment to be subject to the availability of each individual participant. As a result, assignments for the experiment were determined by a partially random block design. In addition to the questionnaire, the pre-game segment involved training and introduction of the scene setter (Kretz, Simpson, & Graham, 2012).

In order to accurately measure the effectiveness of each analytic technique, experiment participants needed to be trained on the various techniques. The Raytheon Grand Challenge researchers recruited participants from the engineering employee population at Raytheon Intelligence, Information, and Services division in Garland, Texas; therefore, most of the participants lacked experience in using the prescribed analytic techniques. The training was also valuable because it established uniform baseline knowledge of the analytic technique for all participants. The introduction of the scene setter in the pre-game experiment segment oriented the participants to the analytic problem. The scene setter was presented in the form of an intelligence report that described the current state of Baghdad. Additionally in the pre-game phase, researchers offered an explanation of the exercise materials and process (Kretz, Simpson, & Graham, 2012).

The “game phases” segment of the analytic exercise marked the actual beginning of the analytic exercise. Participants worked to perform military intelligence analysis through a series of three phases. At the start of each phase, participants were informed of an explosion from an Improvised Explosive Device (IED) near U.S. and Coalition forces. Participants then opened a message set that contained important information related to the event. The message sets contained evidence gathered up to three weeks prior to the attack that included: synthetic phone intercepts, U.S. embassy reports, eyewitness reports, and combat patrol updates. Researchers gave participants the task of reviewing the messages, analyzing the information, and formulating conclusions to specific intelligence questions about the main details of the attack. A checkpoint occurred at the end of each evidence review session, which required participants to record their answers to the intelligence questions regarding the intended target, motivation of the attack, and responsible party. After the third checkpoint, the analytic exercise concluded with a post-game segment (Kretz, Simpson, & Graham, 2012).

The post-game segment involved a questionnaire, performance review, and discussion between the facilitators and participants. The questionnaire was utilized to help researchers gain objective feedback about the difficulty of the problem, the effectiveness of the analytic technique training, the impact of time on completing the task, and the usefulness of the exercise materials. Raytheon researchers utilized the post-game discussion to help them better understand the thought processes of participants throughout the exercise; however, researchers did not take this information into heavy consideration when assessing the data. Raytheon primarily used information collected in response to the intelligence questions in order to calculate the experiment results (Kretz, Simpson, & Graham, 2012).

Raytheon evaluated the participants' intelligence question responses and examined patterns of variation in the judgments participants made throughout the three phases to identify bias in the subjects' analytical decision-making within a specific experimental condition. Raytheon's method to identify bias focused on a preferred explanation for the target of each of the three IED attacks. The evidence pertaining to each attack supported a wide range of possible answers for who was the target of the attack; however, careful evaluation of the evidence encourages a specific explanation for the target of attack for each IED explosion (Kretz, Simpson, & Graham, 2012).

Throughout the exercise, information was introduced in a manner that would likely prompt participants to be influenced by confirmation bias and anchoring effect. The scenario included sufficient evidence to conclude that international forces were the target of all three IED attacks; however, in the final evidence phase new information strongly indicated that the target of the second attack was the leader of the individuals responsible for the IED attacks. If the participant had not been affected by cognitive biases, the introduction of this evidence would have prompted the subject to reevaluate the credibility and likelihood of the anti-coalition hypothesis. If the participant had fallen victim to the cognitive bias of anchoring effect, then he or she would have placed a greater emphasis on the fact that U.S. and Coalition forces were the target of the first IED attack. It is likely that, in participants who remained anchored on the first IED attack against international forces, this anchoring effect would cause confirmation bias to develop and cause the subject to judge evidence in a way that confirmed his or her pre-existing beliefs. As a result of the scenario design, Raytheon expected that participants would have been affected by these cognitive biases and may have ignored, discounted, or created alternative explanations for the detonation report that revealed the attack was motivated by internal IED cell conflict (Kretz, Simpson, & Graham, 2012). Beyond identifying bias exhibited by participants across all experimental conditions, Raytheon evaluated the effectiveness of the three analytic techniques utilized by participants in reducing bias (Kretz, Simpson, & Graham, 2012).

Raytheon calculated for each analytic technique the percentage of participant responses that accurately identified the preferred explanation for the target of each attack. Raytheon hypothesized that unbiased analysis would contain judgments that matched the preferred hypothesis. This hypothesis was based on an assumption by Raytheon that the judgments of participants were primarily based on the exercise data; Raytheon felt comfortable making this assumption because their pre-game questionnaire responses revealed that subjects had little knowledge about the situation in Baghdad, Iraq. Raytheon was also able to identify bias based on the accuracy of responses for hypotheses concerning the target of the attacks because the scenario design was intended to introduce bias within a specific context. Researchers reviewed the responses for each attack over the entire exercise to determine if each tool helped to reduce bias. Once researchers calculated the bias reduction results for each tool, they compared the techniques against each other (Kretz, Simpson, & Graham, 2012). The specific results of Raytheon's pilot study will be discussed later on with a comparison to the results of the Penn State Red Cell Analytics Lab Grand Challenge.

The research results, methodology, data analysis, and limitations are discussed in detail in a co-authored paper presented at the IEEE Conference on Technologies for Homeland Security by Colonel Jacob Graham and Raytheon researchers Donald R. Kretz and B.J Simpson. In this paper, Raytheon identifies a few limitations that occurred due to the nature of their experiment being a pilot study. One of these limitations is the size of their sample; Raytheon had 27 participants for the exercise. Furthermore, the sample was comprised of individuals who lacked intelligence analysis training and experience, as well as subject matter expertise on the security situation and IED problem in Baghdad, Iraq. The non-analyst element of the participants and the little amount of analytic technique training decreases the generalizability of the results to a population of intelligence analysts. After identifying the limitations, Raytheon outlined multiple tasks that they would like to implement in future research. These steps included increasing the scope of checkpoint questions, increasing the breadth of the post-game questions, conducting the experiment with a larger sample size of trained analysts, and changing the experiment procedure to include automation. Specifically, Raytheon sought to automate the data collection and other elements of the experiment design to allow for easier execution of the administrative functions of the exercise (Kretz, Simpson, & Graham, 2012). With Raytheon's permission and partnership, the Red Cell Analytics Lab Grand Challenge team conducted an experiment that implemented these changes, as well as others, to contribute to Raytheon's data analysis and research. Raytheon has conducted several other experiments since this pilot study, as well as sustained its partnership with the College of IST, in an effort to continue studying cognitive biases using an analytic decision game-based approach.

.

# Chapter 4

# Designing the Penn State Red Cell Analytics Lab Grand Challenge

## Institutional Review Board Process

Due to the fact that the Red Cell Analytics Lab Grand Challenge involved a research experiment with human participants, the first step of the project was to complete the training required by the Pennsylvania State University. Our initial research team had four individuals who would be in charge of conducting the experiment, and all four individuals needed to successfully complete the training. Penn State partners with the Collaborative Institutional Training Initiative (CITI) to offer training that meets the standards required by the Penn State Institutional Review Board (IRB). CITI offers multiple IRB required training courses based upon the type of research that will be conducted. The social science training course required participants to complete a series of mandatory modules and one elective module (Office for Research Protections, 2014, February). The IRB training course was very valuable because it outlined the potential problems that can occur during experiments with human participants and the ways to best ensure that researchers accurately and ethically address these problems. Once the training was completed by all researchers, our team began drafting the IRB research proposal.

The drafting of the Institutional Review Board proposal involved extensive planning of the experiment in order to adequately answer all of the questions requested by the Penn State Office for Research Protections. The first step was for our team to solidify the general logistics of the experiment including: the location of the experiment with a description of the facilities available for the research location. Additionally, we were required to record information regarding details of funding and potential conflicts of interest. These sections comprised the beginning of the IRB draft proposal and addressed basic questions pertaining to the study.

The experiment purpose section contained a detailed description of the study and was one of the most important sections of the IRB proposal. This description included background information about the research topic, the rationale for conducting the research, the primary objective of the research, detailed information about the experiment timeframe, criteria for inclusion and exclusion of participants, and the research methodology. The research methodology was the most difficult task to complete because it required our research team to formulate the step-by-step experiment procedures. Specifically, our methodology needed to describe the entire process of how subjects would be asked to participate in the research. Our team developed an eleven step participant involved research procedure, after careful evaluation of the research process conducted by Raytheon and of the adaptations that we wanted to implement for our experiment. Once we solidified the tasks that participants would be asked to complete, our team needed to provide more information about the expected participants.

The Institutional Review Board electronic proposal document asks detailed questions about the number of participants, age range of participants, and participant demographics. The information gathered in this section is an important step to show that researchers are not being discriminatory in their selection of participants. Furthermore, the researchers are required to list steps that will be taken to ensure that subjects are not coerced into participating, or vulnerable to coercion, based on the involvement of professors or employers in the conduction of the research. The participant information section ensures the protection of all Penn State students and faculty, as well as the researchers, in the conduction of and recruitment for a human-subjects experiment.

Our team formulated a recruitment plan, during the process of drafting the IRB proposal. This recruitment plan involved deciding the type of materials to be used, how we would identify and distribute recruitment information to potential participants, and the investigators responsible for approaching and communicating with the potential participants. Additionally, our team evaluated the effectiveness of the recruitment plan in order to ensure that we would be able to attain the required number of study participants. Our team recorded all the details of the recruitment plan, including the explanation for how we would be able to reach our target number of participants, in the IRB proposal. After establishing how our team would obtain participants for the study, we had to review the requirements of participant informed consent.

Our research team decided that all investigators would be responsible for obtaining informed consent from every participant within five minutes of the start of the experiment. At this time, the researchers who obtained consent would also provide participants with a copy of the form to keep for their records. Our team utilized the Penn State University informed consent template, available on the Office for Research Protections website, to guide the creation of our informed consent form (Office for Research Protections, 2014, March). Our team created this document during the IRB drafting process in order to upload the form with the submission of our completed IRB proposal.

The IRB proposal also requires researchers to answer questions regarding compensation and costs to participants. Our study did not impose any additional costs to the participants; however, we planned to offer compensation in the form of extra credit. Our team communicated with various faculty members to determine if they would be willing to offer extra credit to their students who participated in our research study. During our discussions with various professors, our team formulated an alternative assignment for individuals who did not want to participate in the experiment; information about this alternative assignment was provided in our proposal. Although this extra credit adhered to IRB standards, during the actual recruitment period of our research, our research team decided that we would be able to reach our intended number of participants without offering extra credit.

In order to automate the data collection administrative function of our experiment, our research team decided to provide the questionnaires to participants electronically using the Qualtrics web survey system. Qualtrics was used to both collect and store our electronic survey data. Our team was required to provide detailed information about the security and confidentiality of Qualtrics in our IRB proposal. Penn State supports the use of Qualtrics and the Penn State information technology services department provided researchers with Qualtrics user accounts. The IRB proposal requires researchers to discuss the security of commercial service providers in order to evaluate the security of participant information and the risks conducting research and storing research data.

Risk analysis is a large aspect of the IRB proposal. Researchers are required to identify any and all potential risks to participants, as well as explain the likelihood and severity of these risks. Researchers must also have a clear plan for how they will minimize the potential risk to participants. Our research experiment contained no foreseeable risks beyond the risks that participants could incur while using a computer for three hours in an office space setting. To complement the risks, the Institutional Review Board is also interested in the participants' benefits. Our participants benefited by gaining a stronger understanding of structured analytic techniques. Loss of confidentiality of the data is a risk that researchers must also consider and plan to prevent.

Our research procedure involved randomly assigning a non-identifiable number to each participant at the start of the experiment in order to protect the privacy of the participants. A master list linking participants and participant identification numbers was created in the event that researchers needed to contact participants for further information; however, this information was stored in a secure location separate from the electronic data. Furthermore, no personally identifiable information will be shared with Raytheon, nor will it be shared in the event of publication related to the research. Researchers informed participants of the confidentiality of their participation in the experiment, verbally and written in the informed consent form, prior to the start of the exercise.

The final step of the IRB proposal process included uploading documents that would be used in the conduction of the research. For the document upload, our team submitted the questionnaires that would be used to collect data from the participants and the informed consent form. Prior to submitting our proposal, our research team met with IST faculty member, Nick Giacobe, who has expertise in social science research. Dr. Giacobe reviewed our proposal draft and offered guidance for areas of improvement. This information was immensely helpful and beneficial to our research team achieving approval of our proposal from the Institutional Review Board.

Our team had initially completed an IRB application in the expedited research category. After nearly two months, we learned our research fell into the exempt category. This new status required that we change our application; however, the revised application did not take long to complete. Within one week, the Penn State Office for Research Protections approved our IRB research proposal.

Gaining the Institutional Review Board's approval to begin our research was a crucial and rewarding step in the overall research process. For our research team, the IRB proposal draft process was very lengthy, but it prepared us to conduct a higher quality of research. Because our team primarily consisted of students with little research experience, this preparation and thorough experiment review was a very important step in the research process. The IRB process is extremely beneficial because it forces researchers to evaluate all aspects of their research experiment, prior to beginning the conduction of their research. Although the IRB process is tiresome and can delay the intended research timeline, it is a vital step to ensure the protection of human participants and ensure ethical standards in university research.

## Redesign Phase

Due to the nature of the partnership between Raytheon and the College of IST, Raytheon provided our research team with the materials that they had used in their pilot study. These materials included: the pre-game questionnaire, exercise warm-up, acronym list, intelligence questions, intelligence report, facilitator experiment script, analytic technique training materials evidence message sets, checkpoint questionnaires, postgame questionnaire, IBM® Analyst's Notebook association chart, and Google™ Earth kmz file. Raytheon gave our research team full permission to edit the materials. Our team of researchers was able to use a majority of these materials as they were originally provided by Raytheon; these included the exercise warm-up, acronym list, intelligence questions, intelligence report, IBM® Analyst's Notebook association chart, and Google™ Earth kmz file. However, changes in the experimental design also necessitated that researchers adapt the other materials.

Our first adaption took the form of converting the various questionnaires into on-line Qualtrics surveys. During this conversion, researchers added an open-ended question to the pre-game questionnaire to gain more specific information about participants' past experiences in intelligence analysis. Researchers modified the post-game questionnaire to ask participants questions about their travel experiences and educational experiences specific to the College of IST. These questions were added to help researchers gain more insight into the participants' backgrounds, in an attempt to explain possible differences that may have arisen between subjects' performance. As a result of the questionnaires being offered online, as well as researchers offering electronic copies of the message sets, researchers needed to update the facilitator's experiment instruction script. These modifications were relatively minor.

The research team for the Red Cell Analytics Lab Grand Challenge also found it beneficial to update the analytic technique training material tutorials. Colonel Jacob Graham, research advisor on this project, has taught two of the structured analytic techniques featured in this analytic exercise for many years in his classroom. He generously provided some of his classroom instruction material to be used as training materials for the RCAL Grand Challenge. Colonel Graham's expertise in instruction, as well as my past experiences using some of these techniques in the classroom, provided me with the knowledge necessary to accurately teach the fundamental usage of these tools to the participants. After our research team had created our own training materials and modified the experiment design, we conducted a practice experiment introduction and training session for each of the analytic techniques.

Our research partner, Donald R. Kretz, traveled to State College to provide feedback and suggestions on the training tutorials and the experiment modifications. Our research team decided that it would be more beneficial for Mr. Kretz to provide feedback on each of the training sessions, as opposed to witnessing our team running a full-scale practice exercise. The importance of providing an effective training session for each technique motivated this decision. Our team also solicited general exercise recommendations and advice from Mr. Kretz that would guide the pilot conduction of our full-scale exercise. The participants to be included in this practice session were the individuals who were slated to assist me in facilitating the research experiment. During the training material practice session, Colonel Graham identified some limitations of the proposed Information Extraction and Weighting (IEW) technique. In another significant contribution the exercise, Colonel Graham developed a modified approach to the IEW analytic technique.

The Information Extraction and Weighting method that Colonel Graham developed incorporated a sense-making element that did not exist within Raytheon's IEW technique. In the context of this tool, sense-making refers to evaluating each piece of evidence and recording observations that help explain the evidence. Sense-making during IEW is a two-step process; the first step requires an analyst to evaluate the evidence pieces individually. This first judgment represents the analyst's initial judgment, assessment, or inference about the piece of information. In other words, analysts are evaluating the evidence at face value. The second step asks analysts to take into consideration all pieces of information collectively. Analysts formulate alternative explanations and inferences, after considering each piece of information in the context of the entire evidence set. Analysts should conduct sense making in order to understand the importance of individual pieces of information with regard to how they answer key intelligence questions and to how they interact with other significant evidence pieces. The second step of sense-making is also helpful in understanding conflicting pieces of evidence.

After performing sense-making in IEW, analysts weight each piece of evidence. Raytheon had utilized a 10 point scale ranging from marginally important to extremely critical during their pilot study. The RCAL Grand Challenge research team modified the information weighting process to be based on a 3 point scale that assessed the relevance value of a piece of evidence. This relevance value reflects the analyst's assessment as to how relevant a piece of information is to answering the intelligence question being considered. The three point scale labeled evidence as having a high, medium, or low level of relevance to answering the question; this process can be used to determine the potential importance of each piece of evidence in determining the correct solution. Once an analyst has assigned all the pieces of evidence a relevance value, they should sort the key items of evidence, in descending order, according to relevance value. Analysts can use the newly sorted evidence list, placing emphasis on the most relevant pieces of information, when forming key judgments and assessments. Colonel Graham also developed an instruction sheet, scenario, and template to easily train analysts in this new technique. Graham's revision of the IEW technique and creation of training materials were significant contributions to the exercise.

Another significant contribution of the RCAL Grand Challenge was the addition of a fourth experimental condition. As previously mentioned, Raytheon's pilot experiment contained three experimental conditions corresponding to three analytic techniques. Our research team thought it would be valuable to add in a control group condition to the experiment design. Participants in the control group did not have access to any technical structured analytic techniques. Individuals were instructed to utilize their own experiences with analytics to solve the problem. Researchers would evaluate control group data to identify for the presence of cognitive bias. More importantly, the control group experimental condition would be compared against structured analytic techniques to see if the techniques produced a significant reduction in bias. The addition of the control group experimental condition was an important contribution of our research because it helped offer a more comprehensive measurement of the effectiveness of structured analytic techniques in mitigating bias. Prior to including the control group in the actual experiment, a pilot test was conducted on the training session participants.

These participants had no prior knowledge about the outcome of the scenario; therefore they were a viable group on which to test the design of the control group experimental condition. This pilot test included running a full-scale experiment from start to finish. This pilot experiment was extremely beneficial because it helped the researchers practice the experiment procedure, as well as helped to identify potential limitations to the experiment design. The participants of the pilot helped to identify technical problems with the supporting software programs, provide feedback on the proposed control group design, and voice concerns about the time constraints of the experiment. Researchers took into account the feedback from the pilot session and used this information to guide decisions about how to modify various components of the exercise and in particular, the time interval of the game phases. This pilot session was also beneficial to the individuals who participated in the exercise. These participants gained a better understanding of the exercise, which helped them to better serve in their role as research assistant during the facilitation of the analytic exercises. The conduction of the pilot experiment was the final step in designing the Penn State Red Cell Analytics Lab Grand Challenge research experiment.

# Chapter 5

# Experiment Methods

## Participants

Participants were individuals between the ages of 18 and 40 who have an interest in analytics. Researchers recruited exercise participants primarily from the Red Cell Analytics Lab, a student organization within Penn State's College of Information Sciences and Technology. Researchers also advertised the experiment to students at a lecture about careers in the United States Intelligence Community. However, only one of these students participated in the experiment. Majority of the participants were students majoring in Information Sciences and Technology or Security and Risk Analysis. Participants who are upperclassmen in the Security and Risk Analysis major should have had classroom experience with the structured analytic techniques and synthetic intelligence analysis. Furthermore, some participants had military experience; these participants were trained in and exposed to intelligence analysis through that experience.

## Procedures

Researchers conducted the experiment in a series of eight analytic exercise sessions. Each of the experimental conditions was measured on two different days, in order to evaluate three experimental conditions and one control condition. The experimental conditions evaluated one specific structured analytic technique; these techniques included: Link Analysis, Analysis of Competing Hypotheses (ACH), and Information Extraction and Weighting (IEW). The duration of each experiment session was three hours, and each participant participated in only one condition. Participants were assigned to a condition based on the day for which they signed up to participate in the study. Prior to the experiment, participants did not know in which experimental condition they would be participating. Link Analysis, ACH, IEW, and the control group conditions had 13 participants, 10 participants, 11 participants, and 8 participants, respectively.

In each of the eight sessions, one investigator handed each participant a participant identification number as he or she walked into the experiment facility. Researchers then instructed participants to fill out a sheet of paper that listed the assigned participant identification numbers; participants were instructed to provide their name, e-mail address, and telephone number in the row that corresponded with their assigned participant identification number. The investigator will keep the records of the identifiable information separate from the research data; the participant identification lists will be destroyed after statistical analysis is completed. After participants filled out their personal information, researchers pointed them to an assigned computer workstation that corresponded with their participant identification number. Researchers created electronic folders in which to store research data for each individual participant identification number. The facilitators directed participants to work at a specific computer workstation in order to ensure that the data saved on the computers was properly labeled and matched their participant identification number. The specific procedures for the eight experimental sessions involved the following steps.

In order to measure the level of experience of each participant, their familiarity with analytic techniques, and their subject matter knowledge, investigators asked participants to complete the "Pre-Game Background Information Questionnaire". This questionnaire, and all questionnaires used in the experiment, was offered via the Qualtrics online survey platform. Researchers provided a document with electronic links to each of the survey questionnaires on the workstation of each subject to provide participants access to the surveys. Participants then completed an exercise warm-up; this seven question survey was used to determine the amount of prior knowledge that a participant had on the experiment's subject matter and includes a Cognitive Reflection Test (CRT). According to our research partner, the CRT can be used to indicate whether an individual has the predisposition to engage in analysis or if an individual is likely to choose more obvious, but incorrect answers. Researchers can use this information to look for patterns with their exercise performance.

Following the completion of the questionnaire, investigators reviewed the "Exercise Warm-Up" answers with the entire session of participants in order to briefly familiarize all participants with the topic. Our research team also provided participants with a paper copy of the answers in case they needed to refer back to the information during the exercise. The analytic exercise portion officially began after the review of the "Exercise Warm-Up" answers.

To begin the exercise, I explained to the participants that they would be assuming the role of an IED intelligence analyst in a tactical intelligence unit. A brief overview of the available tools was provided followed by an explanation of the analytic technique they would be using. At this time, I provided a training session to the participants demonstrating how to utilize their specific analytic tool. The training session for the control group was less complex than the training procedures for the three structured analytic technique sessions; nevertheless, researchers made every effort to keep the technique instruction time for the control group equal to the amount of time provided for training in the three analytic tool conditions.

In the control condition, the investigators explained that participants could use the suite of Microsoft Office products including: Word, Excel, PowerPoint, and Project. The control group did not have access to any structured analytic techniques. To ensure that there was no confusion among participants, I provided a brief demonstration of how to access the available computer tools to the control group participants. Participants in all of the conditions could also use pens, pencils, highlighters, and scratch paper to support their analyses. After the technique instruction part of the experiment, I provided participants with a verbal explanation of the reference materials.

Reference materials for the experiment included an acronym list, a list of intelligence questions, an intelligence report on the Iraq IED problem, a paper copy of IBM® Analyst's Notebook association chart, and Google™ Earth kmz file with pre-labeled geographic coordinates that pertain to locations in the message data. For the next set of instructions, I read to participants an explanation of the synthetic message data that participants should have used as evidence during the analytic process. A few comments about the scenario were also provided at that time. I inquired as to if participants had any questions before moving on to explaining the procedures of the analytic decision game.

The first procedure explained was that evidence would be given to the participants in three phases and at a specific time. Participants were given a set amount of time to analyze each message set. At the conclusion of each phase allotted for analysis, the participants were asked to save the work of their completed analytic tools into a folder on the desktop of his or her computer. Researchers ensured that every participant had saved his or her work prior to providing the next set of instruction. Facilitators then asked participants to open a specific Qualtrics questionnaire and answer a series of checkpoint questions about their analytic and thought processes. Analysis of a set of messages, followed by completion of checkpoint questions, occurred in each of the three phases.

The analytic decision game concluded after three phases and three subsequent phase checkpoints; at this time, investigators requested that participants complete the Qualtrics "Post-Game Exercise and Demographics" questionnaire. The experiment was concluded as researchers collected all reference materials and messages. After the materials were collected, participants completed one final survey about their opinions related to the exercise. Investigators and participants then had an open discussion about the analytic intelligence exercise. Participants were not required to stay for this discussion, but almost all participants stayed for a portion of the post-exercise discussion. During this discussion, facilitators attempted to answer all questions from the participants; however, researchers were unable to provide information about the preferred answers to the exercise questions. Researchers also recorded participant feedback during the post-exercise discussion to provide insights as to limitations of the research and how the exercise could be improved in the future. At the conclusion of the post-exercise discussion, researchers collected the saved data from the workstations to be stored until analysis was performed.

## Measures

One of the primary questions that my research seeks to determine is if the structured analytic techniques produced a reduction in analytic bias. This study utilized the same measure for bias reduction that Raytheon used in their Grand Challenge experiment. As previously explained in further detail in Chapter 3, Raytheon proposed that bias reduction, in the context of such an experiment, is determined by the accuracy of participant responses. Performance in bias reduction being measured by the percentage of responses that correctly determined the target of each IED attack is based on the assumption that the hypotheses of unbiased analysis would match the preferred attack target explanation. This performance measure can be utilized because researchers designed the scenario to promote confirmation bias and anchoring effects, within a specific context, among the participants. Participants used a self-report measure to record their post-analysis hypotheses for the target of each IED attack.

# Chapter 6

## Results

Before we could evaluate if the use of structured analytic techniques produced a reduction in bias, we needed to determine if analysts exhibited bias in their analysis. We examined participants' responses, as a whole, to the checkpoint questions pertaining to the intended target of each of the three attacks in order to identify the presence of analytic bias. We analyzed the responses and identified that the 42 participant answers fell into nine categories. For each of the three phases, we calculated the percentage of responses that fell into the category that represented the preferred explanation for the target of the attack. Table 2 shows the percentage of responses, out of the total participant responses, for each target hypothesis category. This table also shows which category had the highest number of responses for each attack per evaluation phase, as well as the preferred hypothesis category for each attack.

**Table 2 Percentage of Participant Responses Per Target Hypothesis Category**

| | **Phase 1** | **Phase 2** | | **Phase 3** | |
|---|---|---|---|---|---|
| | **Attack 1** | **Attack 1** | **Attack 2** | **Attack 1** | **Attack 3** |
| **Supplies** | 0.02 | 0.02 | 0.00 | 0.02 | 0.00 |
| **Infrastructure** | 0.07 | 0.05 | 0.05 | 0.05 | 0.00 |
| **Inter-Sect** | 0.12 | 0.12 | 0.07 | 0.07 | 0.02 |
| **U.S. / Coalition** | **0.55*** | **0.69*** | **0.43** | **0.71*** | **0.88*** |
| **Intra-Sect** | 0.00 | 0.00 | 0.05* | 0.07 | 0.10 |
| **Civilians/Property** | 0.10 | 0.02 | 0.00 | 0.00 | 0.00 |
| **Unknown** | 0.05 | 0.02 | 0.07 | 0.05 | 0.00 |
| **No Response** | 0.02 | 0.02 | 0.24 | 0.00 | 0.00 |
| **Multiple Responses** | 0.07 | 0.05 | 0.10 | 0.02 | 0.00 |

***** Indicates preferred target of attack hypothesis
**Indicates highest percentage of responses**

The results show that the majority of participants correctly identified the target of the first and third attacks. As the phases progressed, the percentage of participants who correctly identified the target of the first IED attack increased. The results show that only 5% of the participants, or 2 subjects out of 42 subjects, correctly identified the second attack target as being the result of an internal conflict within the Rashid IED cell. Based on our metric, these results show that the study's participants demonstrated bias in their assessments.

The next objective of our study was to evaluate the effectiveness of the structured analytic techniques in reducing bias. Table 3 shows the percentage of participant responses per analytic tool that correctly identified the target of each attack.

**Table 3 Percentage of Participant Responses, by Analytic Technique, that Correctly Identified the Target of Each Attack**

| | Phase 1 | Phase 2 | | Phase 3 | |
|---|---|---|---|---|---|
| | **Attack 1** | **Attack 1** | **Attack 2** | **Attack 1** | **Attack 3** |
| **Link Analysis** | 0.46 | 0.62 | 0.08 | 0.54 | 0.92 |
| **ACH** | 0.60 | 0.60 | 0.00 | 0.60 | 1.00 |
| **IEW** | 0.55 | 0.64 | 0.09 | 0.82 | 0.64 |
| **Control** | 0.63 | 1.00 | 0.00 | 1.00 | 1.00 |

The control group outperformed the three structured analytic techniques in all three phases in identifying the preferred target hypothesis for the first attack. For phase one, the ACH group performed the next best followed by IEW and Link Analysis, respectively. IEW performed the second best for identifying in phase two the correct target for the first attack; Link Analysis followed IEW, and ACH performed the worst out of the four conditions for attack one in phase two. IEW and Link Analysis outperformed ACH and the control group for accuracy of the second attack responses evaluated in phase two. Neither ACH nor the control group had any participant correctly identify the target hypothesis for the second attack. IEW and Link Analysis each had one participant correctly identify the second attack target. In phase three, the order of performance for attack one, from best to worst, was control group, IEW, ACH, and Link Analysis. ACH and the control group both had 100% of their participants correctly identify the third attack target; Link Analysis performed the third best, and IEW performed the worst for attack three in the third phase. Table 4 summarizes the frequency of each performance ranking for the four techniques.

**Table 4 Performance Rankings for each of the Structured Analytic Techniques**

| | 1st Place | 2nd Place | 3rd Place | 4th Place |
|---|---|---|---|---|
| **Link Analysis** | 0 | 1 | 2 | 2 |
| **ACH** | 1 | 1 | 2 | 1 |
| **IEW** | 1 | 2 | 1 | 1 |
| **Control** | 4 | 0 | 1 | 0 |

The results show that the structured analytic techniques did not significantly aid in reducing bias. The control group outperformed the analytic techniques four times and produced more accurate results for the first and third attacks; however, none of the control group participants correctly identified the second attack target. Participants primarily demonstrated bias in their assessment of this second attack. When comparing the structured analytic techniques, IEW seems to have performed the best because it performed better than the other two tools for three responses. ACH performed better than Link Analysis. However, the performance differences among the tools are not large enough to allow researchers to conclude that one structured analytic technique reduces bias more effectively than the others.

## Comparison of PSU RCAL Grand Challenge and Raytheon Grand Challenge Results

Both the results of the Penn State Red Cell Analytics Lab Grand Challenge and the Raytheon Grand Challenge indicated that experiment participants were impacted by analytic bias. Specifically, the participant responses for the target of the second attack, from both experiments, show that the highest percentage of target hypotheses fell into the U.S./Coalition category. It is likely that these results occurred as a result of the participants assessing the evidence in ways that confirmed the anti-U.S. bias that they had developed from the bias-inducing experiment design. Table 5 shows the breakdown of the percentage of responses for each target hypothesis category that occurred in Raytheon's pilot Grand Challenge experiment.

**Table 5 Raytheon Grand Challenge Percentage of Participant Responses per Target Hypothesis Category. Adapted from "A Game-Based Experimental Protocol for Identifying and Overcoming Judgment Biases in Forensic Decision Analysis" by D. R. Kretz, B. J. Simpson, and J. Graham, 2012, *IEEE Conference on Technologies for Homeland Security (HST).***

| | Phase 1 | Phase 2 | | Phase 3 | |
|---|---|---|---|---|---|
| | Attack 1 | Attack 1 | Attack 2 | Attack 1 | Attack 3 |
| **Intra-Sect** | 0.00 | 0.00 | 0.07* | 0.00 | 0.07 |
| **Inter-Sect** | 0.26 | 0.15 | 0.00 | 0.04 | 0.00 |
| **Local Politician** | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 |
| **U.S. / Coalition** | **0.48*** | **0.67*** | **0.70** | **0.81*** | **0.93*** |
| **Economic** | 0.04 | 0.04 | 0.00 | 0.00 | 0.00 |
| **Sunni Civilians** | 0.04 | 0.00 | 0.04 | 0.00 | 0.00 |
| **Unknown** | 0.15 | 0.15 | 0.19 | 0.15 | 0.00 |

* Indicates preferred target of attack hypothesis
**Indicates highest percentage of responses**

As shown in Table 5, 70% of Raytheon's participant responses identified the U.S/Coalition category as the target of attack two. Only 43% of the RCAL Grand Challenge responses identified this same target for the second attack. Based on the comparison of these two percentages, Raytheon participants may have been more highly impacted by anti-U.S. confirmation bias. It is important to note that RCAL Grand Challenge participants were still highly impacted by bias because only 5% of participants correctly identify the intra-sect category as the target of the attack; however, this bias was distributed across multiple target hypothesis categories.

Table 5 shows that the Raytheon responses that contained bias, in response to the second attack, were split among four categories. In comparison, RCAL Grand Challenge biased responses were split across seven categories. We identified nine categories, in total, that RCAL Grand Challenge participant responses fell into. In contrast, Raytheon categorized its participants' responses into seven target hypothesis categories. In addition to differences among some of the categories, the RCAL categories also included "No Response" and "Multiple Responses" categories. It is important to note that Raytheon's results are based off of a sample size of 27 participants; whereas, RCAL results are based off of a 42 participant sample size. These differences in sample size, as well as differences in analytic experience among the participants, may account for a larger number of categories produced by the RCAL Grand Challenge responses. This difference in the number of categories may also be explained by differences in the time allowed for participants to respond to the checkpoint questions or differences in criteria used by researchers to code the data.

Researchers for both Raytheon and Penn State concluded that experiment results did not indicate a consistent reduction of bias among the structured analytic techniques of Link Analysis, Analysis of Competing Hypotheses, and Information Extraction and Weighting (Kretz, Simpson, & Graham, 2012). Table 6 depicts Raytheon's results for bias reduction.

**Table 6 Raytheon Percentage of Participant Responses, by Analytic Technique, that Correctly Identified the Target of Each Attack. Adapted from "A Game-Based Experimental Protocol for Identifying and Overcoming Judgment Biases in Forensic Decision Analysis" by D. R. Kretz, B. J. Simpson, and J. Graham, 2012, *IEEE Conference on Technologies for Homeland Security (HST)*.**

| | Phase 1 | Phase 2 | | Phase 3 | |
|---|---|---|---|---|---|
| | Attack 1 | Attack 1 | Attack 2 | Attack 1 | Attack 3 |
| **Link Analysis** | 0.33 | 0.56 | 0.22 | 0.67 | 1.00 |
| **ACH** | 0.44 | 0.78 | 0.00 | 0.89 | 0.89 |
| **IEW** | 0.67 | 0.67 | 0.00 | 0.89 | 0.89 |

Raytheon results showed that the techniques of ACH and IEW were relatively equal in their effectiveness in reducing bias. For example, the two techniques performed equally in the third phase and alternated in being the best performing tool in phases one and two. Furthermore, the table shows that Link Analysis had the best performance for the second and third attacks. Raytheon results also indicate that Link Analysis was the only technique that reduced some of the anti-U.S. confirmation bias exhibited by the participants in their analyses. Similarly in the RCAL study results, Link Analysis and IEW each aided one participant in overcoming analytic bias concerning the target of the second attack. However, in contrast between Raytheon's results and RCAL's results, the other techniques outperformed Link Analysis when considering performance over the entire experiment.

Results of the Raytheon Grand Challenge and the RCAL Grand Challenge confirm the findings of prior research that confirmation bias exists within the process of intelligence analysis. Furthermore, researchers of both experiments conducted separate data analyses and reached the same conclusion: the structured analytic techniques of Link Analysis, ACH, and IEW do not significantly reduce analytic bias. The inconsistencies in analytic technique performance within the results of each experiment, as well as the differences in technique performance in comparison of the two experiments, are the primary pieces of evidence to support this conclusion.

# Chapter 7

# Discussion

## Potential Limitations

The first potential limitation of my study involves participant application of the structured analytic techniques to the intelligence problem. Although researchers highly encouraged the use of the assigned structured analytic technique within each of the experiment conditions, it is very difficult to ensure that all participants utilized the analytic method specified. Researcher observations and participant discussion from our experiment revealed that some participants used the analytic technique sparingly or as an analytic afterthought. As opposed to using the tool for his or her primary method of analysis, participants would develop their own process to conduct analysis. This limitation may impact the results of my study because the results of each experiment condition may not accurately reflect the effectiveness of the specified analytic technique. Furthermore, this limitation expands beyond simple participant choice to discontinue the use of the analytic method.

The study cannot ensure that participants are accurately using the tool. For example in our experiment, facilitators observed that some participants utilized the tool in customized, unique ways. The results for these participants may not reflect the effectiveness of the structured analytic technique in reducing bias; but rather, the results may reflect the effectiveness of the participant's individual process or method of analysis. Researchers need to consider if improper use of the analytic method should still be included in measuring the effectiveness of the techniques.

The understanding of analytic techniques could also be a limitation of the study. The participants in our experiment had different levels of experience in using analytic techniques. While some had previous practice using their randomly assigned technique, many of the participants in our experiment were unfamiliar with their condition's analytic process. This unfamiliarity ranged from participants being unsure of the purpose of the technique to subjects having a very limited knowledge of the ways to utilize the method. Although all participants went through a training tutorial on the basic elements of the technique, a lack of adequate understanding of the analytic method may have impacted the results. The techniques may not have shown a reduction in bias because the participants did not fully understand how to best utilize the technique to conduct analysis. More experienced users in the techniques may produce results that more accurately represent the effectiveness of the specified analytic method.

In addition to a fundamental lack of understanding of the analytic technique, some participants may have also had issues with the software products the experiment utilized to help facilitate the use of these analytic methods. For example, our experiment promoted the use of IBM® Analyst's Notebook to perform Link Analysis. The participants may have fully understood the ways to perform Link Analysis; however, they may have had some trouble in performing the functions as a result of using the software program as opposed to simply building a link analysis diagram by hand. Although researchers did not discourage participants from performing the technique without the aid of the electronic tool, the technique tutorial sessions incorporated the specific software tools. The encouragement of the use of specific software programs to perform the analytic techniques is a potential limitation of the study.

The results may have been potentially impacted by the time constraints of the experiment. Post-experiment discussion with the participants revealed that many participants felt that they needed more time to conduct the analysis. It is highly likely that the majority of answers that fell into the "No Response" category resulted from the game phase time constraints. In order to maintain consistency across all experiment conditions, researchers had to follow strict time frames for the message analysis and checkpoint completion times of the exercise. Researchers used feedback from the pilot experiment to determine time frames for each phase that would allow adequate, but limited time for analysis. Time constraints were also established to provide an added element of pressure on the participants in an effort to mimic the feeling of pressure experienced by analysts in a real-world setting. However, the participant responses may have been more reflective of the time constraints than the use of the structured analytic technique, which could potentially impact the results of the study in measuring the effectiveness of these techniques.

Review of the participant responses shows that some of the participants may have been confused by the context of the checkpoint questions. Specifically, some participants seemed confused in their responses as to which IED attack the question was referencing. Misunderstanding may have resulted in participants incorrectly identifying the target of the attack, which would increase the level of bias measured in the experiment.

Another potential limitation could be the sample size for each experiment condition. Due to the difficulty of participant availability for a three-hour experiment, participant assignment was based on the day for which participants signed up. The experiment condition was still random because conditions were assigned to a specific day prior to subjects signing up for the exercise and participants were unaware of which technique corresponded with the day for which they signed up. Based on this method, researchers could not ensure that all experiment conditions had the same amount of participants. The variation in experiment condition size may have impacted the measure of performance of the experiment conditions in reducing analytic bias.

The most significant potential limitation of the study is the generalizability of the results beyond my specific research experiment. There is the potential that these results are only representative of the precise factors of time, setting, and participant experience of this specific experiment. Due to the nature of using a fictional scenario, the results of this experiment may not be generalizable to the larger field of professional intelligence analysis. The pressure of real-world intelligence analysis simply cannot be replicated in the lab environment. Intelligence analysis has the potential to impact an entire nation, and consequences from mistakes in this analysis can include loss of human life. The high stakes of intelligence analysis will be extremely difficult, if not impossible, to mimic in a sanitized experimental study.

Furthermore, there is no uniform standard of time allotted for intelligence analysis in the real-world. The amount of time an analyst spends on a given problem can be influenced by a number of factors including, among other things, the level of intelligence assessment (i.e. strategic, tactical, or operational), amount of information available, recipient of the analytic product, priority level of the assessment, and real-time events. The time constraints of the experiment may limit the generalizability of the results to the larger intelligence analysis field, which functions according to dynamic analytic time frames.

The results of this sample population may also not be generalizable to the larger target population of experienced intelligence analysts. Characteristics of my sample population, with consideration to training on the use of structured analytic techniques, experience conducting intelligence analysis, and personal background experiences, may be very different from the characteristics of my target population. Consequently, these differences have the potential to threaten the external validity of my experiment.

## Recommendations for Future Research

Future research can be conducted to address the limitations of my study. One minor improvement includes rewriting the checkpoint questions to increase participant clarity about the context of questions about specific attacks. My recommendation to improve the clarity of checkpoint questions is to specifically state in which phase the attack happened, as opposed to referring to the “first IED attack”. This improvement should sufficiently address the limitation of participant confusion regarding the context of the intelligence questions. Additionally, I suggest that researchers expand the third phase checkpoint to include questions about the target of attack two. Evidence from phase three reveals that the IED cell leader’s mobile phone detonated the IED that killed him in the second attack. Based on the introduction of this new evidence, there is a chance that participants overcame, in phase three, the anti-U.S. bias they demonstrated, in phase two, towards the target of the second attack. A more comprehensive phase three checkpoint questionnaire may be additionally useful in helping researchers to identify bias and to determine the effectiveness of the structured techniques in reducing this bias.

To address the limitations pertaining to participant application of a given structured analytic technique to an intelligence problem, it is recommended that researchers conduct pre-experiment training sessions for each technique to a degree that ensures adequate familiarity with that technique. Time constraints of the tutorial provided during the experiment do not allow participants to become adequately prepared or comfortable with using the techniques. With the introduction of a pre-experiment training that includes instruction of the technique and a chance for participants to practice using the technique, it is likely that the results would be more representative of the actual effectiveness of the analytic techniques and be more representative of real-world analysis. Researchers could still provide a technique tutorial at the start of the experiment; however, the tutorial could be used to simply refresh participant knowledge of the various functions of the assigned structured analytic technique. An additional consideration is the emphasis on analytic process over the use of a specified software program; that is to say, the construction of hand drawn Link Analysis charts as opposed to the use of IBM® Analyst's Notebook software.

Future researcher should consider conducting these experiments with participant samples directly from the target population of professional intelligence analysts. This recommendation would allow for the results to be more generalizable to the population. Even though it is likely impossible to mimic the pressure of real-world intelligence analysis in an experimental study, this recommendation would still be beneficial because the participants would have a better understanding of the high stakes associated with intelligence analysis.

Intelligence analysis often takes place in a team setting; therefore, future research should consider facilitating this experiment with participants working in teams to conduct analysis and decision-making. The results of this research would likely be more generalizable to the larger field of intelligence analysis. The first goal of this research should be to identify and measure the extent that judgment bias exists in the analytic products produced by teams. The second research objective should be to study the ways that teams influence the effectiveness of structured analytic techniques in mitigating analytic bias. It would be interesting to compare the results of an experiment that studied team-based use of analytic methods with the results of my study.

There are a few modifications to the exercise design that I would recommend for future research. The process of intelligence analysis is often predictive in nature or focused on assessing the current state. In order to capture these aspects of intelligence analysis, researchers should introduce the attacks in a delayed time frame. For example, as opposed to starting each phase with the attack, I think it would be valuable to have participants conduct on-going analysis prior to the introduction of the attacks. Additional checkpoint questions could be added to assess the participants' judgments for on-going analysis. Intelligence analysis expands beyond forensic analysis, which is why I think researchers should modify the experiment design to have participants analyze evidence for indications of early warning topics. This modification would add an element of predictive intelligence to the experiment. Furthermore, the spontaneous introduction of attacks and the disruption to current evidence analysis would make the scenario more reflective of a real-world tactical intelligence environment.

Another recommendation to the exercise design would be incorporating a more interactive and comprehensive all-source, multi-intelligence (MULTI-INT) evidence set. For example, evidence could include synthetic recorded phone intercepts, satellite images, and video surveillance. Additionally, the evidence set could include some pieces of unprocessed data. The creation of these evidence pieces would require extensive work and time; however, it would make the experiment much more realistic of the real-world tactical intelligence environment. Additionally, participant focus would likely increase as a result of having data collected by multiple types of intelligence included in the evidence set.

The final recommendation for future research would be to conduct the experiment over a period of time that is longer than three hours. I acknowledge that it is difficult to get subjects to participate in a three hour exercise; however, the overall experiment would be much more realistic, and likely produce more accurate results, if conducted over a longer period of time. A majority of my recommendations for future research would require a longer experiment time frame in order to be successively implemented. Based on the impact of time on the successful implementation of my research recommendations and the potential impact of time on the results of the experiment, researchers should place high priority on this recommendation.

# Chapter 8

# Conclusion

The results, limitations, and recommendations of my study, and the Raytheon Grand Challenge pilot experiment, support the conclusion that there is much more research that needs to be conducted on the effectiveness of structured analytic techniques in mitigating bias exhibited during the intelligence analysis process. In an even more frustrating conclusion, I recognize that even if research experiments using analytic decision games are able to prove that structured analytic techniques are an effective bias reduction method, these tools cannot be effectively tested in the real world. In the field of intelligence analysis in the real world, hindsight is often the only way to identify biased intelligence assessments. Research studies will continue to be limited in their generalizability to the real world because experiments are unable to fully represent the conditions that exist during the process of professional intelligence analysis. The best response to this conclusion is to hope that researchers will continue to improve experiments and address, with the exception of the element of pressure felt by intelligence analysts, all limitations to current research.

Some of the structured analytic techniques that we examined in this study are currently used by intelligence analysts in the real world. If research continues to find that these techniques do not produce a significant reduction in bias, we must question whether or not organizations should continue to advocate for their use during the intelligence process. The results of my study showed that a higher percentage of participants who did not use the structured analytic techniques correctly identified the target of the attacks than participants who had used the analytic procedures. Although the control group was unable to overcome their anti-U.S. bias for the second attack, the results as a whole indicate that participants in the non-technique condition demonstrated a lesser amount of bias in their assessments. If more studies replicate the results of my study, should we continue to promote the use of these specific structured analytic techniques? Furthermore, even though participants who did not utilize structured analytic techniques outperformed the other participants, they still exhibited bias in their assessments. These results support the conclusion that researchers must continue to investigate various analytic techniques and methods to determine the most effective procedure to reduce bias during the analytic process.

Discussions about historical examples of analytic bias, as well as review of research findings, can also aid in reducing bias because analysts will become more cognizant of the problem. If an individual is aware of their potential for bias then they can make a conscious effort to check their analysis for biased assessments. As an individual who is interested in analytics, it is my belief that the most significant contribution of my research, on a personal level, is my heightened awareness and understanding of analytic bias.

As a result of the culmination of the entire research process, I gained a comprehensive understanding of the problem of judgment bias exhibited during analysis. Research into the influence of bias, most strongly demonstrated by the failure in 2003 of analysts in the United States Intelligence Community to accurately assess the Iraqi nuclear weapons program, increased my knowledge of the extent of the problem and emphasized the importance of overcoming bias. The review of literature available on previous research experiments and an evaluation of Raytheon's Grand Challenge experiment helped me to better understand the strengths and weaknesses of current research on analytic bias. However, the actual facilitation of my own research experiment provided to be the most personally beneficial element of my research.

My understanding of the various functions of the structured analytic techniques of Link Analysis, Analysis of Competing Hypotheses, and Information Extraction and Weighing significantly increased as a result of this experiment. Although I had exposure to Link Analysis and ACH in the classroom, my knowledge primarily consisted of how to utilize the various functions of the tools. I am much more confident in my understanding of the ways these techniques can aid analysis after facilitating this experiment. The process of teaching these methods to the participants helped me to better understand the conceptual aspects of the procedures and identify the common aspects of confusion about these techniques. Through participant observation, I was able to identify multiple ways of using the technique to approach solving the analytic problems. This observation offered me insight into different thought processes and contributed to my personal knowledgebase of methods that can be used to conduct analysis. As a result of the experiment, my level of understanding of the specified structured analytic techniques is much deeper.

The process of data analysis provided further insight into the different ways that individuals can make connections and approach solving an analytic problem. It was very interesting to see the different conclusions that individuals drew from analysis of the same evidence set. These results showed me the various ways that evidence can be interpreted to support multiple hypotheses, which emphasized the importance of thoroughly considering all plausible hypotheses when conducting analysis. These lessons are a valuable contribution to my personal interest in analytics and will improve my ability to assess evidence to form accurate, reasonable analytic judgments. The knowledge gained from the entirety of this research experience will significantly strengthen the quality of my decision-making process.

The contributions of my research efforts to the field of literature on judgment bias will hopefully motivate fellow researchers to continue exploring the topic of analytic bias in an effort to determine the most effective techniques to mitigate such bias. It is my further hope that the results of this study can assist individuals in improving the quality of their analyses for producing sound analytic judgments. The process of intelligence analysis influences life-or-death situations; thus, the consequences of analytic bias cannot be underestimated.

# BIBLIOGRAPHY

Belasco, A. (2011, March 29). *The cost of Iraq, Afghanistan, and other global War on Terror operations since 9/11* (Congressional Report No. RL33110). Washington DC: Library of Congress Congressional Research Service. Retrieved from: http://www.fas.org/sgp/crs/natsec/RL33110.pdf.

Cook, M., & Smallman, H. (2008). Human factors of the confirmation bias in intelligence analysis: decision support from graphical evidence landscapes. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(5), 745-754. DOI:10.1518/001872008X354183

Costs of War Project, Watson Institute for International Studies, Brown University. (2013). *Iraq: 10 years after invasion.* Retrieved from: http://www.costsofwar.org/iraq-10-years-after-invasion

Crawford, N. C. (2013). *Civilian death and injury in the Iraq War, 2003-2013.* Retrieved from: http://www.costsofwar.org/sites/default/files/articles/15/attachments/Iraqciv2013.pdf

Crawford, N. C. (2013). *U.S. costs of wars through 2013: $3.1 trillion and counting: Summary of costs for the U.S. wars in Iraq, Afghanistan and Pakistan*. Retrieved from: http://costsofwar.org/sites/all/themes/costsofwar/images/Us_Costs_of_Wars.pdf

Hall, D., & Graham, J. (2011). *Statement of work for user-centric analytics research and collaboration between Raytheon IIS and PSU IST*. Unpublished manuscript, College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA.

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology, 80*(4), 557-571. DOI:10.1037/0022-3514.80.4.557

Kretz, D.R.; Simpson, B.J.; Graham, J. (2012). A game-based experimental protocol for identifying and overcoming judgment biases in forensic decision analysis. Paper presented at *IEEE Conference on Technologies for Homeland Security (HST),* Boston, Massachusetts, 439-444.

Lehner, P., Adelman, L., Cheikes, B., & Brown, M. (2008). Confirmation bias in complex analyses. *IEEE Transactions On Systems, Man, and Cybernetics, 38*(3), 584 - 592. DOI: 10.1109/TSMCA.2008.918634

Lehner, P., Adelman, L., DiStasio, R., Erie, M., Mittel, J., & Olson, S. (2009). Confirmation bias in the analysis of remote sensing data. *IEEE Transactions On Systems, Man, and Cybernetics, 39*(1), 218-226. DOI: 10.1109/TSMCA.2008.2006372

Office for Research Protections, The Pennsylvania State University. (2014, February 20). Training on the protection of human research participants (IRB required). Retrieved from http://www.research.psu.edu/training/irb-basic-training

Office for Research Protections, The Pennsylvania State University. (2014, March 18). Guidelines, templates and sample consent forms. Retrieved from http://www.research.psu.edu/orp/humans/conducting-study/preparing-new-application/informed-consent/guidelines-templates-sample-icfs

The Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction. (2005). *Report to the president of the United States*. Retrieved from: http://hdl.handle.net/2027/umn.31951d02402919g

Torreon, B. S. (2012, December 28). *U.S. periods of war and dates of current conflicts* (Congressional Report No. RS21405). Washington DC: Library of Congress Congressional Research Service. Retrieved from: http://www.fas.org/sgp/crs/natsec/RS21405.pdf

U.S. Department of Defense. (2014). *Casualty status*. Retrieved from: http://www.defense.gov/news/casualty.pdf

Wheeler, P. R., & Arunachalam, V. (2008). The effects of decision aid design on the information search strategies and confirmation bias of tax professionals. *Behavioral Research in Accounting, 20*(1), 131-145. Retrieved from http://search.proquest.com/docview/203296927?accountid=13158

# ACADEMIC VITA

Melissa Sirera
127 North Sparks St. #18
State College, PA 16801
mmsirera@gmail.com

## Education

The Pennsylvania State University, University Park, PA

B.S. Security and Risk Analysis– Intelligence Analysis & Modeling
B.S. Global and International Studies—concentration on the Levant
Minor: Arabic Language; May 2014
Minor: Information Sciences and Technology; May 2014
Minor: Middle East Studies – Regional Focus: Middle East; May 2014
IA Certificate—Information Systems Security Professionals (NSTISSI No. 4011)

Council of International Educational Exchange: The University of Jordan, Amman, Jordan Arabic Language and Culture Program; Spring 2012

CompTIA Security+ Certification; May 2013

Federal CyberCorps Scholarship for Service (2012-2014)

Strategic and Global Security Program (2012-2014)

## Honors and Awards

College of Information Sciences and Technology Student Marshal for Spring 2014 Commencement Ceremony
Pennsylvania State University Evan Pugh Senior Scholar Award (2014)
Pennsylvania State University Evan Pugh Junior Scholar Award (2013)
Carey Lynne DeMoss Memorial Award for Excellence in Building a Global Perspective (2012)
Pennsylvania State University Freshmen President's Award (2011)
Schreyer Honors College Academic Excellence Scholarship (2010-2014)
College of Information Sciences and Technology Schreyer Student Scholarship (2010-2014)
Dean's List (Fall 2010, Spring 2011, Fall 2011, Spring 2012, Fall 2012, Spring 2013, Fall 2013)

## Work/Research Experience

Research Assistant, Strategic Intelligence Research Program, PSU Applied Research Lab (Summer 2013)
Red Cell Analytics Lab, Network-Centric Cognition and Information Fusion (Fall 2012 to present)
Learning Assistant, Pennsylvania State University College of IST (Fall 2012)
Business Systems Analyst, Technology Intern, PNC Financial Services Group, Inc. (Summer 2012)

## Activities

Gamma Tau Phi Honors Society (2013-2014)
Red Cell Analytics Lab Operations Officer (2013-2014)
Information Assurance club member (2012-2013)
Women in IST (WIST) club member (2010-2012)
Penn State Interfraternity Council/Panhellenic 46 Hour Dance Marathon

Independent fundraising dancer couple (2012-2013)
OPPerations committee member (2010-2011)