APPROPRIATE FACTORS TO CONSIDER WHEN ASSESSING ANALYTIC CONFIDENCE IN INTELLIGENCE ANALYSIS

JOSHUA J. PETERSON

A Thesis

Submitted to the Faculty of Mercyhurst College

In Partial Fulfillment of the Requirements for

The Degree of

MASTER OF SCIENCE
IN
APPLIED INTELLIGENCE

DEPARTMENT OF INTELLIGENCE STUDIES
MERCYHURST COLLEGE
ERIE, PENNSYLVANIA
MARCH 2008

DEPARTMENT OF INTELLIGENCE STUDIES
MERCYHURST COLLEGE
ERIE, PENNSYLVANIA

APPROPRIATE FACTORS TO CONSIDER WHEN ASSESSING ANALYTIC CONFIDENCE IN INTELLIGENCE ANALYSIS

A Thesis
Submitted to the Faculty of Mercyhurst College
In Partial Fulfillment of the Requirements for
The Degree of

MASTER OF SCIENCE
IN
APPLIED INTELLIGENCE

Submitted By:

JOSHUA J. PETERSON

Certificate of Approval:

_____
Kristan J. Wheaton
Assistant Professor
Department of Intelligence Studies

_____
James Breckenridge
Chairman/Assistant Professor
Department of Intelligence Studies

_____
Barbara A. Behan
Vice President
Office of Academic Affairs

MARCH 2008

To be uncertain is to be uncomfortable, but to be certain is to be ridiculous.
-Chinese Proverb

Without a humble but *reasonable* confidence in your own powers you cannot be successful or happy.
~Norman Vincent Peale

DEDICATION

This work is dedicated to:

-Tim, Gail, Tracy, and my grandparents for their continued support and confidence in me, no matter what factors it is based on.

-Dr. Tamara Wilkins and Sgt. Josh Lego for inspiring me as two of the most erudite people I have ever met, taking the time to encourage and challenge me during my time at MSU.

-Ryan Hartmann for being my own personal Tenzing Norgay over the last 13 years

ACKNOWLEDGEMENTS

I would like to acknowledge my thesis advisor and primary reader, Kristan Wheaton, for his guidance and thought provoking discussions over the course of this work.

I would like to thank Professor Hemangini Deshmukh for her instruction and assistance in performing the statistical analysis on this work, it was a great help.

I would also like to thank Mr. Richards Heuer for making some of his work and personal correspondence available to me in furtherance of my research.

Finally, I would like to acknowledge Jen Lee and Mike Lyden for making their work available to me, and their advice throughout the process.

ABSTRACT OF THE THESIS

Appropriate Factors To Consider When Assessing Analytic Confidence In Intelligence Analysis

By

Joshua J. Peterson

Master of Science in Applied Intelligence

Mercyhurst College, 2007

Professor Kristan J. Wheaton, Chair

[Analytic confidence in intelligence analysis is a topic on which very little specific research has been done, yet has an incredible impact on the United States Intelligence Community's ability to accurately inform policymakers. This study examines what little literature exists on analytic confidence in intelligence forecasting, in addition to discussing relevant studies from the social sciences, and those factors found to have an impact on confidence in decision-making. After examining each of these factors, an experiment was conducted to test the hypothesis that these factors are appropriate factors upon which to rate analytic confidence in intelligence analysis. The findings of the experiment suggest that those factors are indeed valid ones to consider when assessing analytic confidence, though more research is recommended to make the experiment's results more robust. This thesis concludes with recommendations for future research, and this author's assertion of his own method to assess analytic confidence as discussed in the work.]

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER ONE:

INTRODUCTION

"It's a slam dunk," is probably the most infamous statement of analytic confidence in recent history. Those words, spoken by then-director of Central Intelligence George Tenet, expressed his certainty over finding weapons of mass destruction in Iraq.[1] Ironically, while history has proven this sure bet to have been wrong, it is a rare example of expressing analytic confidence within the Intelligence Community. To eliminate this rarity of expressing analytic confidence, the Intelligence Reform and Terrorist Prevention Act, signed into law by President George W. Bush on December 17, 2004, explicitly called for intelligence analysts to express analytic confidence in their analyses:

> "[Reviews of finished intelligence products] should include whether the product or products concerned were based on all sources of available intelligence, properly describe the quality and reliability of underlying sources, properly caveat and express uncertainties or confidence in analytic judgments, properly distinguish between underlying intelligence and the assumptions and judgments of analyses; and incorporate, where appropriate, alternative analyses."[2]

This directive has been realized, in part, as evidenced in the July 2007 National Intelligence Estimate (NIE).[3] Contained within this NIE, and those following it, is a short section which attempts to explain the difference between analytic judgments and

---

[1]William Branigin, "CIA Director Tenet Resigns," 06/03/04, Washington Post, http://www.washingtonpost.com/wp-dyn/articles/A12296-2004Jun3.html (accessed May 14, 2007).

[2] United States Government, *Intelligence Reform and Terrorism Prevention Act of 2004,* December 17, 2004, http://www.nctc.gov/docs/pl108_458.pdf.

[3]National Intelligence Council, "The Terrorist Threat to the US Homeland," July, 2007, Office of the Director of National Intelligence, http://www.dni.gov/press_releases/20070717_release.pdf (accessed August 31, 2007).

analytic confidence levels. Following this is an even briefer overview of how analytic confidence is rated and that it is based on "the scope and quality of information supporting our judgments."[4] While this brief mention is certainly a step toward incorporating analytic confidence ratings in every analytic work, it is confusing to read, awkward in its application, and does not leave the reader with a clear understanding of the NIE's analysts' confidence in the very same report. Clearly more changes need to be made.

Expressing analytic confidence in intelligence analysis would have two enormous benefits. First, decision makers will be able to see the analyst's confidence in his/her own assessment, possibly even question them about their confidence, and thereby make a more informed choice as a result. Second, greater accountability would be achieved through analytic confidence because it would provide a type of audit record which could be reviewed by superiors and oversight committees in the future. This would leave less ambiguity in what went into the decision maker's mind when acting on a piece of finished intelligence.

Despite these benefits, there is one hurdle to overcome before analytic confidence can be utilized to its full potential: What should and should not be taken into account when formulating an expression of analytic confidence? The United States Intelligence Community bases its confidence ratings on only two factors: source reliability and the difficulty of the question.[5] While this thesis supports source reliability as an important factor to consider in analytic confidence, it will go on to suggest that it is only *one of*

---

[4]Ibid.

[5]National Intelligence Council, The Terrorist Threat to the US Homeland," http://www.dni.gov/press_releases/20070717_release.pdf.

many factors to contemplate when formulating an analytic confidence rating. After all, with the current system of focusing exclusively on source reliability, what happens when two highly reliable sources' directly conflict? In relying only on source reliability analysts set the stage for a poorly calibrated confidence rating, and thus may affect a key decision from policymakers or military leaders.

This thesis will argue that despite the relative dearth of studies focused specifically on confidence in intelligence analysis, studies and experiments in other fields concerning accurate measures of confidence in decision making shed light on what are appropriate measures to consider when formulating a well calibrated analytic confidence rating in intelligence analysis.

The purpose of this study is to determine precisely what should and should not affect an intelligence analyst's construction of a confidence rating in his/her analysis, and to demonstrate that those factors are indeed true measures of what analytic confidence should be based upon. Doing so will go a step beyond the requirements laid out in the Intelligence Reform and Terrorist Prevention Act of 2004 by moving away from *intuitive* methods of estimating analytic confidence and moving toward a *method* of calculating/assessing analytic confidence which is based upon research and evidence applied to the analytic tradecraft.

Based upon that, the following hypotheses were formulated: First, that the concepts identified in the proceeding literature review are appropriate factors to consider when assessing analytic confidence in intelligence analysis. Second, that these factors when compiled and tested together in an experiment will have an increasing or decreasing affect on experimental subjects' analytic confidence, with regard to the factors

being manipulated so as to appropriately increase or decrease analytic confidence in high confidence and low confidence groups respectively. Third, that when subjects perform a routine analysis without any discussion of analytic confidence or the factors identified in this thesis, their analytic confidence will vary greatly within that group (control group).

The result of this will be a more accurate and standardized reflection of analytic confidence in intelligence analysis. As one study on confidence and uncertainty in decision making has warned:

> People tend to be overconfident and not well calibrated in their evaluation of their response, it is essential for decision makers to be cautious when making critical decisions.

CHAPTER 2:

LITERATURE REVIEW

Before examining the literature pertaining to analytic confidence it is important to cover some definitions. To begin with, there is a need to differentiate analytic confidence from the actual analysis and the words of estimative probability contained therein. For example that 'it is highly likely there are weapons of mass destruction in Iraq.' While the likelihood of that analysis being true is quite high in the eyes of the analyst, it is based on the available evidence. Analytic confidence is the analyst's rating of his/her confidence in the information available, the way they have gone about analyzing it, and the analysis itself, a measure that goes beyond simply what the evidence points to and conveys how comfortable or confident the analyst is in his/her forecasted analyses. The estimate is based on evidence; analytic confidence is based on the analyst's confidence in the analysis, and the process and inputs by which is was created. Though it can be confusing, in realm of intelligence analysis both are important and can make all the difference in the world.

Analytic confidence is also to some extent separate from any rating of source reliability from which the analysis was constructed, though source reliability is certainly one of many factors to be considered when assessing analytic confidence. It is possible to have a finished piece of analysis with a very high level of source reliability and yet an extremely low level of analytic confidence. An example of this would be when all of the evidence gathered points to one outcome, say that Iraq has weapons of mass destruction;

however the analyst's confidence in the report could be quite low.[6] The reason for this lowered level of analytic confidence could be, as this thesis will outline, due to the complexity of the issue, lack of knowledge/experience on the part of the analyst, time pressure, or not having used any structured methods in the analysis.

It is also prudent to clarify the difference between psychological confidence and analytic confidence. Psychological confidence can be likened to a 'feeling' about a particular subject. It is not based on any scientific or verifiable system; instead it is entirely subjective and varies greatly according to biases and heuristics. An example would be a student feeling confident their collegiate hockey team will win the national championship after going to only 1 game and knowing nothing about the sport of hockey. *Analytic* confidence differs because it is not based on a 'feeling'; instead it is a more 'legitimate' rating of one's confidence in a particular analysis and the analytic process by which it was formed. In the example, the student's analytic confidence in his/her hockey forecast should be quite low due to factors such as a lack of understanding of the sport, seeing only one game, the statistical chances of any team winning a national championship in a particular year, along with a whole array of other factors.

The Origins of Analytic Confidence

Efforts to better understand analytic confidence are rooted in studies on decision making in the field of cognitive psychology, which is the school of psychology that

[6]Schrage, What Percent Is 'Slam Dunk'?," http://www.washingtonpost.com/wp-dyn/articles/A37115-2005Feb19.html.

examines internal mental processes such as problem solving, memory, and language.[7] Cognitive psychology, a relatively new sub-field within psychology as a whole, began to take shape in earnest in the 1950's with the actual term "cognitive psychology" coined by Ulric Niesser in 1967.[8] Within the cognitive sub-field, judgment and decision making is a branch of interest and research. It is here, within the realm of psychologists and academics that analytic confidence has its psychological roots.

As one might expect, there is yet another level of sub-groups within the judgment and decision making concentration, however only one of these truly relates to confidence in decision making and analysis. Psychological decision theory, based primarily on the work of Amos Tversky and Daniel Kahneman[9], focuses on the decisions people make and how they go about making them (a vast and intricate field in and of itself). Their work, beginning in the 1970's, supplemented by the work of Baruch Fischhoff[10] from the 1980's onward, is the foundation on which this specialized field been built. As will be seen in the latter sections of this literature review, other social scientists, including many cognitive psychologists, have taken further steps pursuing these research interests.

It is important to note that while the monumental work of scholars like Tversky and Fischhoff were focused on decision making, it was not directed toward the concept of

---

[7][Wapedia], "Cognitive psychology," 11/15/2007, Wapedia, http://wapedia.mobi/en/Cognitive_psychology (accessed November 20, 2007).

[8]Dr. Michael R.P. Dougherty, "What is Cognitive Psychology?," University of Maryland, http://64.233.167.104/search?q=cache:IS9FuGJ8xDcJ:www.bsos.umd.edu/psyc/dougherty/classes/Psyc341/history%2520of%2520cognitive%2520psychology.ppt+http://www.bsos.umd.edu/psyc/dougherty/classes/Psyc341/history%2520of%2520cognitive%2520psychology.ppt&hl=en&ct=clnk&cd=1&gl=us (accessed November 20, 2007).

[9]Amapedia, "Kahneman & Tversky," 02/23/2006, Amazon.com, http://amapedia.amazon.com/view/Kahneman+&+Tversky/id=109612 (accessed November 20, 2007).

[10]Carnegie Mellon University, "Baruch Fischhoff," 2005, Carnegie Mellon University, http://sds.hss.cmu.edu/src/faculty/fischhoff.php (accessed November 20, 2007).

analytic confidence. Instead, it sought to understand the inner processes and reasons by which humans' psychological confidence in their decision making develops. Thus, the main thrust of discussing social scientific thought and research is not to speak of its direct bearing on analytic confidence, but instead to explore what factors have been found to be accurate measures of well calibrated confidence in decision making. It is through this lens that research in decision making applies to this thesis on analytic confidence in intelligence analysis.

In a more narrow scope, the history of analytic confidence as a concept in the field of intelligence analysis has only recently begun to come into the public limelight. Though there can be little doubt that there have been past internal efforts by intelligence agencies to incorporate some form of analytic confidence into their analyses, the documentation of these initiatives are not available to the public. However a few articles have been made public on the website of the Central Intelligence Agency discussing analytic confidence in their estimates over approximately the past five years.[11]

Much of the discussion within the CIA's Center for the Study of Intelligence is centered on improving intelligence analysis and the expressing the associated confidence in it, has centered on psychological confidence in analyst's estimates. Efforts to better convey the likelihood of forecasted events can be traced all the way back to a pioneer of the CIA's analyst school: Sherman Kent. Kent promoted the use of "Words of Estimative Probability" which, as stated before, are used in expressing what the evidence in the analysis points to. These words have previously been confused as expressing

[11]Central Intelligence Agency, "Center for the Study of Intelligence," 09/17/2007, Central Intelligence Agency, http://https://www.cia.gov/library/center-for-the-study-of-intelligence/index.html (accessed November 20, 2007).

confidence in the analysis, when in reality they are expressions of likelihood. Thus, while many analysts attending the CIA's school, named after Kent, may think they know what analytic confidence is and how to express it, it is quite possible that instead they are expressing a level or degree of likelihood in their forecasts, though no details on how to measure levels/degrees of likelihood or how they are derived, are given.

However, in spite of analytic confidence not being directly addressed until early this century, there are some very brief mentions as early as 1964 in a recently unclassified "classic piece" written by Kent himself.[12] In it Kent tells analysts studying particularly difficult or vague topics that:

> Obviously no one expects you to be wholly accurate or very confident of your findings. But you people are after all the experts, and it would be too bad if I had to go to others for this stuff who know far less about it than you.
>
> Yet on many occasions a writer will feel uncomfortable--and justifiably so--with a bare "It is likely that. . . ." Such a bald statement is seemingly more confident than the situation would warrant. The writer will feel something akin to a compulsion towards modesty and a drive to soften the "likely" by introducing it with a "we believe" or "we estimate."[13]

In the first quote Kent, probably without even realizing it, is alluding to the need for an avenue by which analysts can express their confidence on subject matter they may not feel comfortable working with. His comment, which sounds almost like a coach's pre-game pep talk, is an effort to persuade analysts to take a stand on their analyses and realize that they will not be right about everything all the time. Similarly, the context of the second quote also highlights the need to include some form of confidence expression,

---

[12]Sherman Kent, "Words of Estimative Probability," Fall 1964, Central Intelligence Agency, http://https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/sherman-kent-and-the-board-of-national-estimates-collected-essays/6words.html#ft10 (accessed November 20, 2007).

[13]Ibid.

which wouldn't necessarily "soften" the estimate but would allow analysts to convey their uneasiness in their forecasts. These references to analytic confidence, while not direct, foreshadowed discussion four decades later by analysts at CIA.

In 2003 Jack Davis, a scholar at the Kent Research Center, published "Analytic Professionalism and the Policymaking Process," an article written in a question and answer format discussing many facets of analysis.[14] In the seventh question, Davis mentions "analysts' confidence levels" and what recourse should be taken if a policymaker erroneously attributes the wrong level of confidence to an analyst's forecast.[15] Though Davis uses the term "analysts' confidence levels," the ensuing paragraphs strongly indicate he was in fact speaking to certainty levels in forecasting not analytic confidence.

Davis' article is mentioned for two reasons. First, it shows the degree to which even scholars at the Center for Intelligence Studies seem to be confused about the difference between psychological and analytic confidence. Although Davis used the term "confidence intervals" what he was really referring to was likelihood of the estimative judgment on the part of the analyst. Essentially, Davis was discussing what to do if a policymaker misrepresents the forecasted likelihood of an event in an estimate produced by an analyst, yet he erroneously refers to it as "confidence levels." Second, this mentioning of analytic confidence brings up an important point if taken literally and Davis' proceeding discussion is ignored. If expressions of analytic confidence are included in intelligence analysis, the possibility of those "confidence levels" being

[14]Jack Davis, "Analytic Professionalism and the Policymaking Process: Q&A on a Challenging Relationship," October, 2003, Central Intelligence Agency, http://https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/docs/v02n4p.htm (accessed November 20, 2007).

[15]Ibid.

misrepresented will also exist. Though perhaps not of the same dire importance as misrepresentation of the estimative judgment itself, imagine if the CIA's Iraq WMD team had told George Tenet they had extremely low confidence in their estimate on Iraq's WMD program. Then, instead of echoing his analysts' confidence level Tenet tells the President the CIA thinks it is a "slam dunk." The implications of such a misrepresentation are colossal. Jack Davis asked an extremely important question in his article, he just answered it in a different context.

In December, 2005 Jeffery Cooper wrote an article for the Center for the Study of Intelligence on improving intelligence analysis, which briefly mentioned analytic confidence in its fourth chapter.[16] Discussing methods of communicating both estimative and analytic confidence, Cooper writes:

> Communicating complex judgments and degrees of confidence in those judgments is best done through conversation among the parties, which demands different mechanisms than simple dissemination of "facts." If the mechanisms for interaction with the users of intelligence are designed only to support the provision of individual pieces of evidence rather than to engage both parties in an extended conversation in which ambiguity and subtlety can be communicated, it is unlikely that either party will be satisfied with these interactions.[17]

It is very important to note that here Cooper differentiates between estimative judgments and confidence in those judgments. This quote demonstrates that analysts and those at CSI are beginning to incorporate expressions of confidence in their work, though Cooper's suggestion for conversation may not be wholly practical to the intelligence community. Furthermore it may be inferred that there is still some confusion about the

[16]Cooper, Jeffrey R., "Curing Analytic Pathologies: Pathways to Improved Intelligence Analysis," December, 2005, Central Intelligence Agency, http://https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/curing-analytic-pathologies-pathways-to-improved-intelligence-analysis-1/chapter_4.htm (accessed November 20, 2007).

[17]Ibid.

difference in estimative versus analytic judgment and thus Cooper's suggestion for a dialogue between an analyst and policymaker may be a step toward its clarification. It is in Cooper's dialogue that may demonstrate expressing analytic confidence's true worth. That is an analyst's being able to discuss their reservations and uneasiness about their confidence in their forecast with the policymaker and clarify any questions or misunderstandings between them.

These internal efforts coupled with the Intelligence Reform and Terrorism Prevention Act of 2004 that have slowly carved out a place for expressing analytic confidence in intelligence analyses. Unfortunately, neither the scholars at the Center for the Study of Intelligence, nor the authors of the IC's NIEs, have really clarified what are appropriate factors on which to formulate a well calibrated expression of analytic confidence. Additionally, within the IC as a whole there is no accepted universal way to convey/quantify analytic confidence which only adds to the confusion surrounding the concept.

Although the idea of inserting a formal indicator of analytic confidence in intelligence analysis has entered into the public realm of discussion as a result of recent 'intelligence failures' surrounding the terrorist attacks on September 11, 2001 and Iraq's weapons of mass destruction, efforts to implement this concept have been in the works for some time. Regardless of the exact date when the idea surfaced, the fact that it has not been implemented with any overarching success indicates either an unwillingness to incorporate this concept in analyses on the part of the Intelligence Community, or a lack of ability on the part of the analysts. However, with the passing of the Intelligence Reform and Terrorism Prevention Act of 2004, conveying analytic confidence in some

manner in each final piece of intelligence analysis is now something every analyst has been mandated to do by the US Congress and the President.[18]

One way to express such confidence is a numerical representation, set on a 1-10 scale, accompanying each piece of analysis the analyst produces. While literature within the US Intelligence Community is sparse of this subject, the field of decision-making psychology is lush with research on effective methods of communicating confidence. In fact, in her doctoral dissertation in psychology, Caroline Wesson writes, "While people appear to have a preference for receiving numerical probabilities, as communicators they often feel more comfortable using verbal phrases."[19] Wesson then goes on to explain numerical representations of confidence are typically more precise than their verbal counterparts, then noting the argument against numerical values as carrying the illusion of precision which may suggest the probability of an event is measurable, which in reality it may not be.[20] While Wesson's extremely detailed research, covering many other aspects of decision-making psychology not relevant to this work, leans toward expressing confidence in numeric terms, this is not the only method used.

Verbal or non-numeric expressions of confidence are found, most notably in intelligence, in NIEs. Expressions labeled as high, medium, and low confidence are used by the NIEs' authors to express their confidence in the estimates, which at least according to Wesson sacrifices much of the precision attached to numeric ratings. Regardless of

---

[18]United State Government, "Intelligence Reform and Terrorism Prevention Act of 2004,," 12/17/04, http://www.nctc.gov/docs/pl108_458.pdf.

[19]Wesson, Caroline J., "The Communication and Influence of Confidence and Uncertainty," November, 2005, University of Leicester, http://www.le.ac.uk/pc/bdp5/Cari's%20Thesis.pdf (accessed May 15, 2007).

[20]Ibid.

which method, verbal or numeric, is selected for use, it is imperative to *actually express them*. Interestingly, while four NIE's were made public in the years between the Iraq WMD NIE and the most recently released NIE on Iran's nuclear program, all of which contained sections describing how the Intelligence Community defined analytic confidence and how it was rated, there was not a single actual mention of analytic confidence in the estimative portion of the NIEs![21] The first recent public NIE to actually put analytic confidence into practice was the NIE on Iraq's WMD programs, which contained 8 statements of analytic confidence.[22] Following that NIE, not a single expression of analytic confidence was included in the next 4 public NIEs, despite all of them having sections detailing and defining analytic confidence. With these reports coming from the National Intelligence Council, it can be inferred that the exclusion of statements of analytic confidence in the NIEs between the Iraq WMD NIE and the Iranian nuclear program NIE was not accidental. This was clearly a conscious choice, to leave out analytic confidence expressions, yet keep the section detailing what analytic confidence is and how it is rated. While it is possible that most of the readers of these 4 NIEs did not really care about the lack of analytic confidence ratings contained in them, it is hard to believe that no one in the United State Intelligence Community noticed how careless this looked. Then comes the NIE on Iran's nuclear weapon's program, released in late 2007.[23] In this NIE, almost as if trying to make up for a lack of them in past

[21]Kristan Wheaton, "Part 5 -- Enough Exposition! Let's Get Down To It…," January 8, 2008, Kristan Wheaton, http://sourcesandmethods.blogspot.com/ (accessed January 8, 2008).

[22]Ibid.

[23]National Intelligence Council, "Iran: Nuclear Intentions and Capabilities," 12/03/2007, Office of the Director of National Intelligence, http://www.dni.gov/press_releases/20071203_release.pdf (accessed January 8, 2008).

NIE's, there were 19 statements of analytic confidence, in only 31 sentences of analysis.[24] Not only does this seem excessive when compared to the total number of estimative statements in the document, but it is also worth mentioning that the majority of the statements of confidence were 'high confidence.' In fact, there was only one lone 'low confidence' statement in the entire document.[25] Certainly the US Intelligence Community is world-class, but that many statements of high confidence bring back to mind the many studies of people's tendency toward overconfidence.[26]

Unfortunately, despite the foreseeable benefits toward implementing analytic confidence measures into intelligence analysis, there has been very little published research focusing directly on confidence as it relates to intelligence forecasting. Information can, however, be gleaned from studies and experiments focusing on accurate confidence gauges and decision making in cognitive psychology and other fields within the social sciences. In examining these studies in decision making psychology this literature review will highlight what their results showed regarding valid factors affecting confidence. The impetus behind this being that as factors which affect confidence calibration are discussed, a list of factors analysts should consider when formulating their own analytic confidence expressions will be derived.

---

[24]Wheaton, Part 5 -- Enough Exposition! Let's Get Down To It…," http://sourcesandmethods.blogspot.com/.

[25]National Intelligence Council, Iran: Nuclear Intentions and Capabilities," http://www.dni.gov/press_releases/20071203_release.pdf.

[26]"Overconfidence," 2007, http://overconfidence.behaviouralfinance.net/ (accessed March 2, 2008).

This literature review will examine the few previous works relating to analytic confidence in intelligence analysis, and the findings of various experiments in social science fields attesting to what are appropriate measures to assess in analytic confidence.

Confidence In Intelligence Analysis

Perhaps the most telling reason for the lack of quantitative research done on analytic confidence in intelligence analysis is, as Richards Heuer puts it, "it is difficult to judge over confidence in a probabilistic environment."[27] He then continues:

> When we [CIA analysts] report that the Shah of Iran will "probably" remain in power, we are at the same time implying that "possibly" he may not. When the Shah then falls, was our estimate wrong? Or were we even overconfident in our judgment?[28]

Heuer's point is well put; the nature of what intelligence analysts forecast is inherently more complex and multi-faceted than the simple decisions people make each day, on which there is a wealth of literature in the social sciences. Moreover, when denial, deception, misinformation, and disinformation are factored into an analyst's job, judging the appropriate amount of confidence in reports becomes even more tenuous. Heuer goes on to say that while certain general tests have shown analysts within the intelligence community to be overconfident, he disagrees with such findings based on a number of reasons. One such reason is what he terms 'chain inference,' taking what is the most likely outcome and assuming it is true, then using that assumption to base other inferences.[29] This strategy reduces the large number of probabilities and options an

---

[27]Heuer, Richards J., "Are Analysts Overconfident?," *Unpublished Manuscript* (1980).

[28]Ibid.

[29]Ibid.

analyst faces, but statistically sets up the analyst to appear overconfident. Additionally, Heuer cites contradictory information as another pitfall which causes intelligence analysts to only appear overconfident, when really they are simply trying to, "make inferences from conflicting evidence."[30] In his paper, Heuer says that he does not feel intelligence analysts are as overconfident as some social science tests maintain, and that, "in dealing with probabilistic events, one cannot draw any such conclusion from any single case."[31]

Interestingly, one study has done just that, in his graduate thesis Mike Lyden looked specifically at the estimative accuracy of NIEs when terms such as 'likely' or 'certain' were used.[32] Surprisingly, Lyden found that, when NIEs make estimative statements using words like "likely," they are correct approximately 77% of the time -- not too bad considering the complex issues often dealt with in such documents.[33] In fact CIA Director Mike Hayden concurs on this point when talking about how 'good' the CIA is [on a 1-10 scale] saying:

> "The first thing you've got to understand is, eight and nine aren't on our scale. OK? If it's up at eight or nine, it's generally not the business of intelligence. I mean, intelligence works in a range of things that are inherently ambiguous. And even when we're at the top of our game, it's very, very rare that we can give certitude to a policymaker. And so, one of the things that I would try to do - I am trying to do - is to inform both the public at large and others within the government that, as good as we might be, 10 certainty with regards to our judgments, that's never going to be achieved."[34]

---

[30]Ibid.

[31]Ibid.

[32]Michael Lyden, "The Efficacy of Accelerated Analysis in Strategic-Level Estimative Judgments," (Master's Thesis, Mercyhurst College, May 2007).

[33]Ibid.

[34]Central Intelligence Agency, "Transcript of C-SPAN Interview With CIA Director," April 17, 2007, Central Intelligence Agency/CSPAN, http://https://www.cia.gov/news-information/press-releases-statements/press-release-archive-2007/transcript-of-c-span-interview-with-cia-director.html (accessed March 2, 2008).

Hayden and Lyden both make an excellent point on the ambiguity of intelligence work, however Lyden's second finding is much more surprising. He found that: NIE estimates using words of certainty like "is," "will," and "has" are only correct approximately 57% of the time. That is a 20% difference in the accuracy of estimate statements, and it is counterintuitive to say the least.[35] That the NIE's authors are wrong more often when using words of certainty, than when using words of likelihood, is troubling and does not instill much confidence in many of the NIEs' forecasts. While it is true that there are not ratings of analytic confidence attached to most, if not all, of those estimates, it would be reasonable to conclude that the NIEs' authors were likely to have been more confident in what they were so *certain* about than what they thought was *likely*.

Looking back at what Heuer said previously, it might be unfair to say the intelligence community is overconfident based on one or two wayward estimates, Lyden's study of a sample of 100 estimates contained in 10 NIEs, out of a population of approximately 960 estimates contained in 96 NIEs, spanning from the 1950s to the present, presents clear evidence that the United States Intelligence Community has long experienced problems in confidence calibration.[36]

Richards Heuer also speaks of analytic confidence, though briefly, in his book Psychology of Intelligence Analysis, when discussing how well calibrated analysts are with their confidence judgments.[37] Calibration is a measure of forecasting precision

---

[35]Ibid.

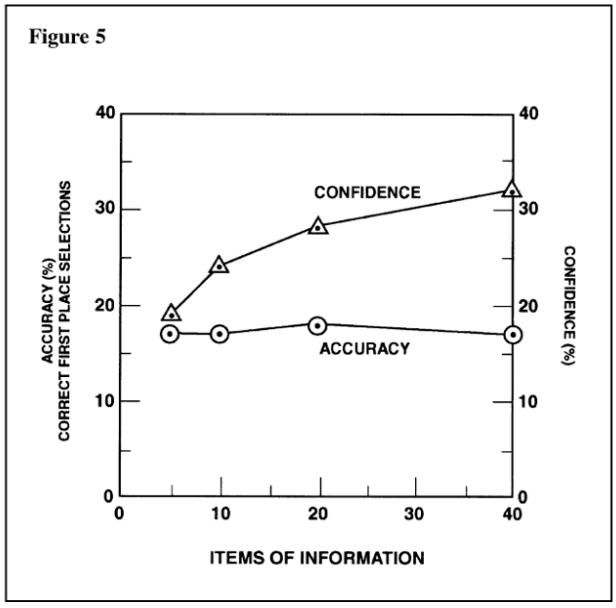
[36]Ibid.

[37]Heuer, Richards J., *Psychology of Intelligence Analysis* [book on-line] (Langley, Virginia: Central Intelligence Agency Center for the Study of Intelligence, 1999); available from http://http://www.au.af.mil/au/awc/awcgate/psych-intel/index.html; Internet.

comparing the confidence in the forecasted event happening, to the frequency of that event.[38] The experiment described in Heuer's book outlines a series of confidence tests involving horse handicappers. The experiment went as follows:

> Eight experienced horserace handicappers were shown a list of 88 variables found on a typical past-performance chart--for example, the weight to be carried; the percentage of races in which horse finished first, second, or third during the previous year; the jockey's record; and the number of days since the horse's last race. Each handicapper was asked to identify, first, what he considered to be the five most important items of information--those he would wish to use to handicap a race if he were limited to only five items of information per horse. Each was then asked to select the 10, 20, and 40 most important variables he would use if limited to those levels of information. Each handicapper was [then] given the data in increments of the 5, 10, 20 and 40 variables he had judged to be most useful. Thus, he predicted each race four times--once with each of the four different levels of information. For each prediction, each handicapper assigned a value from 0 to 100 percent to indicate degree of confidence in the accuracy of his prediction.[39]

Figure 2.1



The results of this were steadily increasing levels of handicapper confidence in their forecasts as more and more evidence was provided, however the accuracy of their forecasts remained virtually unchanged throughout all 4 predictions. (See Figure 2.1) Interestingly, this experiment closely parallels

[38]Yuval, Shahar, "Confidence, Forecasting, Knowledge, and Calibration," http://www.ise.bgu.ac.il/courses/dm/Lectures/lecture11.ppt (accessed May 14, 2007).

[39]Heuer, Richards J., *Psychology of Intelligence Analysis* [book on-line].

intelligence analysis due to the uncertainty and lack of information experienced by the handicappers, especially in the early rounds. Thus, the addition of more information when making an analysis should not necessarily equate to a corresponding increase in analytic confidence.

Confidence in Decision Making

Studies and experiments within the social sciences have proved time and again that people are generally quite poor at matching accuracy and confidence calibration in decision making.[40] People tend to be overconfident in their decisions, thus the calibration between the decisions they make and the accuracy of them is generally low. Thus, relying on peoples' intuitive sense of confidence calibration, in decision making or intelligence analysis, will not produce accurate confidence assessments. Fischhoff and MacGregor struck down earlier assertions that simply asking people to list reasons why their answers might be wrong did little to diminish peoples' natural overconfidence in their decisions.[41] However, they did not offer any suggestions of their own on how to improve confidence calibration.

In place of relying on intuition, numerous studies have focused on what factors promote or adversely effect decision making quality and associated confidence in those decisions. Although there may be situations when the findings of these studies do not

---

[40]Shradha Tibrewal, and John Poertner, "Confidence and uncertainty in casework decisions: The supervisor's role," School of Social Work, University of Illinois at Urbana-Champaign, http://cfrcwww.social.uiuc.edu/pubs/Pdf.files/confidence.pdf (accessed May 15, 2007).

[41]Baruch Fischhoff, and Don MacGregor, "Subjective Confidence in Forecasts," *Journal of Forecasting* 1 1982 [journal on-line]; available from http://http://stinet.dtic.mil/oai/oai?&verb=getRecord&metadataPrefix=html&identifier=ADA109730; Internet.

apply, generally speaking the ensuing sections discuss factors which have been found to promote accurate decision making and as a result should improve confidence calibration. By applying the findings of these experiments, mostly undertaken in the field of psychology, to the concept of what are and are not appropriate measures to be considered in constructing an analytic confidence estimate, intuitive reasoning will be replaced by a method which should yield better confidence calibration.

The factors discussed in the remaining portion of this literature review were included in this work based on a number of criteria. Most importantly, these factors were found to legitimately affect accuracy in decision-making, and thus when present in intelligence analysis should lead to increased, and hopefully more well calibrated analytic confidence. Factors not meeting this first criterion, such as having analysts/subjects list reasons why they could be wrong in their forecast, were not included as many studies have found this practice to have little effect on proper confidence calibration. In addition other factors/exercises found to have little effect on proper confidence calibration include increasing the amount of information available (see above discussion on Heuer's horse handicappers), instruction on inherent tendencies toward overconfidence and consideration of whom the analysis is for. Second, the factors identified are relatively agreed upon in their affect on confidence calibration. True, some of the factors such 'the level of analyst collaboration' have dissenters, which are noted, however as a whole they are agreed upon as valid findings. Finally, these factors were selected because they are readily applied in analytic confidence in intelligence analysis. While other factors may have been found to affect confidence in decision making, such as the amount of information at hand, who the analysis is for, or the amount the analysis differs from the

status quo, they inject a great deal of ambiguity into situation. Not knowing the total amount of information possible, who the estimate is being prepared for, and what the 'status quo' is, would only serve to muck up the analyst's calibration further, if the answers to these questions are ever possible.

Thus, while the factors found in the ensuing review are not all encompassing, the research conducted has found them to be relevant, applicable, and feasible in the general scope of analytic confidence in intelligence analysis. It is on these criteria that the following factors have been selected for discussion and extrapolation into the realm of intelligence analysis.

Subject Matter Expertise

Logic suggests that the more knowledge an analyst has on a given subject, the better his analysis and confidence calibration on the subject will be. However this logical inference was proven incorrect, to some extent, by Philip Tetlock in his book Expert Political Judgment. After compiling results from an extensive 20 year experiment focusing on forecasting accuracy and confidence calibration, Tetlock counter intuitively concludes that subject matter experts are only slightly better at calibrating confidence in their forecasting than dilettantes are. It is important to note that the 'experts' in Tetlock's study were 284 participants with an average of 12.2 years relevant work experience, 52% of which had doctoral degrees, and 96% of which had postgraduate training. The dilettantes consisted of the same group of people, however they were classified as dilettantes when forecasting outside their area of expertise. Both of these groups scored

much higher in forecasting and confidence calibration than did undergraduate students.[42] In fact, despite relatively poor scoring by all forecasters, experts and dilettantes scored higher than undergraduate students throughout the book.

A study in decision making by Tsai, Klayman, and Hastie furthers this point. They found that "in general, experts seem to be better calibrated and less overconfident than novices."[43] The results of their study also confirmed Tetlock's finding that experts' confidence is not perfectly calibrated, only an improvement over non-experts.

Tetlock's work is noteworthy in relation to intelligence analysis due to its utilization of real world situations and events. Participants in his experiment were asked to forecast such events as macroeconomic policies in Latin America, South African elections in 1988, and the demise of the Soviet Union.[44] Clearly, these types of complicated forecasts are experienced in intelligence analysis frequently.

By using complicated statistical analyses, Tetlock comes to the conclusion that subject matter expertise really isn't a panacea to forecasting accuracy and confidence calibration, though it is better to have more knowledge and experience than very little. Essentially, someone with a higher education is likely to be a better forecaster than a high school drop out, as Tetlock's comparison of experts and dilettantes with undergrads shows. He does however interject a word of caution with these results, "as expertise rises, we should therefore expect confidence in forecasts to rise faster, far faster, than

---

[42]Tetlock, Philip E., *Expert Political Judgment* (Princeton, New Jersey: Princeton University Press, 2005).

[43]Claire Tsai, Joshua Klayman, and Reid Hastie, "Effects of Amount of Information on Judgment Accuracy and Confidence," 2006, University of Chicago, http://home.uchicago.edu/~iwentsai/TsaiEtAl_Confidence.pdf (accessed May 16, 2007).

[44]Ibid.

forecast accuracy."[45] Later in his book he reaffirms this point saying, "Beyond a stark minimum, subject matter expertise in world politics translates less into forecasting accuracy than it does into overconfidence."[46] The concept that beyond a certain point, the accumulation of extensive subject-specific knowledge may actually decrease the overarching accuracy of the analysis in some cases has not yet been recognized by the Intelligence Community as a whole.

Evidencing this point is a report recently published by the Center for International and Security Studies at the University of Maryland in which the authors make consistent reference to, "individuals currently entering the analytic workforce will be the seasoned analysts of 2020."[47] Strikingly, the analysts entering the workforce in 2006-2007 will have 13 years of experience in analysis by the time the community recognizes them as 'seasoned,' which is very similar to the experience level of Tetlock's experts. It is undoubtedly true that not all of these analysts will be focused solely on one specific subject, thereby developing a deep expertise. However it is interesting to note that Tetlock's dilettantes scored nearly identical to the experts in forecasting and confidence calibration, with little experience in that particular field.

Furthering this point, Kevin O'Connell, the Director of Defense Group Incorporated's Center for Intelligence Research and Analysis (CIRA) has said that analysts with only a few years experience lack the "context about the world" needed for

---

[45]Ibid.

[46]Tetlock, Philip E., *Expert Political Judgment*.

[47]Lahneman, William J., *The Future of Intelligence Analysis* [book on-line] (Volume 1 Final Report; University of Maryland, 2006, accessed 14 May 2007); available from http://www.cissm.umd.edu/papers/files/future_intel_analysis_final_report1.pdf; Internet.

accurate analysis.[48] Though O'Connell does not give a specific number of years needed to develop thesis skills, he does go on to detail CIRA's use of subject matter experts outside of the Intelligence Community in producing intelligence analyses.

It is not the intent of this thesis to argue that expert opinion does not have a place in intelligence analysis or that it can be replaced by those with less knowledge and experience. Instead, the point made about subject matter expertise as it relates to analytic confidence is that subject matter expertise is only beneficial up to a certain point in many cases, and that beyond that it may actually be detrimental to well calibrated confidence assessments. Similar to Heuer's horse handicappers, the amount of knowledge or intelligence an analyst has does not necessarily correlate to the amount of confidence the analyst should have in his analysis. Ideally an analyst with post graduate level training, as Tetlock's experts and dilettantes had, with a few years of experience would produce the most well calibrated confidence assessments. Unfortunately, at the present time there have been no studies done on what the optimum level of knowledge or experience on a given topic will yield the best forecasts and confidence assessments.

Time Pressure And Analytic Confidence

The effects of time pressure on forecasting accuracy, and the ensuing level of analytic confidence attached to it, is wholly relevant to the fast paced world of the intelligence community. A study of decision-making done by the Israeli Air Force concluded that, "time pressure usually, but not always, impaired performance," a finding which a number of other researchers have verified through a differing array of

[48]Sebastian Abbot, "CIRA and the Business of Revolutionizing Intelligence Analysis," Good Harbor Report, http://www.goodharborreport.com/node/555 (accessed May 14, 2007).

experiments.[49] Similarly, Speier found that information overload, as it relates to time pressure, had an adverse effect on the quality/accuracy of decisions made by subjects in her study. [50] Speier however, went a step further to suggest a model for the relationship between time pressure and decision making performance:

> Similar to the saturation point in information overload, an inverted U-shaped curve is also used to explain decision performance when experiencing time pressure. Moderate levels of time pressure result in accelerated information processing where the decision maker often makes high quality decisions in less time than when time is unlimited. However, when time constraints become more severe, decision accuracy deteriorates as decision makers either reduce their examination of information or consciously choose not to use some information.[51]

Mann and Tan took this concept one step further in their study on the *perception* of time pressure and decision making ability. In that study two groups of people were forced to make complex decisions within the same amount of time. However, one group was led to believe they would have to hurry to complete the tasks, whereas the other group was led to believe that had plenty of time to complete the work.[52] The results of the experiment suggested that the mere perception of time pressure may be as detrimental as actual time pressure, due to the cognitive anxiety brought on by such a perception.[53]

---

[49]Niv Ahituv, Magid Igbaria, and Aviem Sella, "The Effects of Time Pressure and Completeness of Information on Decision Making," *Journal of Management Information Systems* 15, no. 2 fall 1998 [journal on-line]; available from http://jmis.bentley.edu/articles/v15_n2_p153/index.html; Internet; accessed May 15, 2007.

[50]Cheri Speier, *Using Aggregated Data Under Time Pressure: A Mechanism for Coping with Information Overload* [book on-line] (University of Oklahoma, Michael F. Price College of Business, 1998, accessed 15 May 2007), 1-10; available from http://csdl2.computer.org/comp/proceedings/hicss/1998/8236/02/82360004.pdf; Internet.

[51]Ibid.

[52]Leon Mann, and Charlotte Tan, "The Hassled Decision Maker: The Effects of Perceived Time Pressure on Information Processing in Decision Making," *Australian Journal of Management* 12/02/93 [journal on-line]; available from http://www.agsm.unsw.edu.au/eajm/9312/pdf/mann.pdf; Internet; accessed 15 May 2007.

[53]Ibid.

Mann and Tan suggested three plausible reasons for this erosion of decision making accuracy. It is possible that the anxiety of time pressure "motivates a reduction in information search and reliance on simple decision rules."[54] However resentment at being hurried through the task is possible, as is subjects' disrupting their own concentration by continuously checking the clock to see how much time remained.[55]

In a study focusing on other effects time pressure has on decision making, Lee and Dry found that subjects' confidence decreased due to simply having to make more decisions within a given time period.[56] Subjects participating in the experiment had to choose between two doors, while being given advice of varying accuracies, then pick a door and express their confidence in having made the right decision (psychological confidence).[57] They describe the findings:

> There are, however two surprising regularities in the data. The first is that mean confidence on no advice trials decreases across the six conditions. This shows that people become less confident in their guessed decisions when they have to make them relatively often. This is not consistent with equating confidence and accuracy, because the guesses are (on average) equally accurate across the conditions. The second surprise is that mean confidence on advice trials does not increase across the six conditions. It is relatively constant, and perhaps even shows a slight inverted U-shape. The lack of increase shows that people do not become more confident in those decisions based on advice as the accuracy of that advice improves.[58]

It is important to point out that Lee and Dry varied both the accuracy and the frequency of the advice given in their experiment. However, from these findings it can

---

[54]Ibid.

[55]Ibid.

[56]Michael D. Lee, and Matthew J. Dry, "Decision Making and Confidence Given Uncertain Advice," *Cognitive Science* 30 03/08/06 [journal on-line]; available from http://www.socsci.uci.edu/~mdlee/trust_lee_dry.pdf; Internet; accessed 15 May 2007.

[57]Ibid.

[58]Ibid.

be surmised that multiple analyses produced within a short timeframe by a single analyst will generally have a negative effect on the calibration of the analyst's confidence to his/her forecast. Moreover, the complexity of the forecast/decision, in this case merely picking between two doors, leads one to believe that with more complicated analyses the effect may be exacerbated. As time pressure, or the perception thereof, increases, peoples' ability to make accurate decisions decreases. These findings suggest that numerous, quickly put together or rushed intelligence analyses will be less accurate as a whole than those produced under conditions of relatively low pressure. Although time constraint is unavoidable in the Intelligence Community, its detrimental effect on the quality of analyses produced, and its concordant effect on analytic confidence are important to consider.

There is some disagreement on this point however, in his thesis Mike Lyden speaks of a concept called accelerated analysis.[59] This concept or method is essentially one in which the analyst(s) rapidly produce analyses or products in a short amount of time, versus the more traditional method of having a great deal of time to work with. Lyden makes the point that this type of accelerated method/practice is capable of producing estimates which rival NIEs in terms of accuracy and nuance.[60] While I agree with his findings I feel Lyden's general thrust on accelerated analysis carries two large assumptions. First, he says accelerated analysis is analogous to a washing machine in that it is iterative and "utilizes a constant feedback loop" with the decisionmaker, which is unfortunately not always possible in analysis.[61] Second, his comparison and discussion

[59]Lyden, The Efficacy of Accelerated Analysis in Strategic-Level Estimative Judgments."

[60]Ibid.

[61]Ibid.

of nuance in the same section does not give weight to the possibility that classified forms of the NIEs he mentions could actually contain much more nuance. Despite these two criticisms, I feel that Lyden would agree that rushing analysts too much will result in a worse product than having ample time (even for accelerated analysis), and that without a effective feedback loop the iterative nature of accelerated analysis is stopped in its tracks.

An overarching comment regarding time pressure on decision making and analytic confidence is that an ideal analysis and analytic confidence rating would take place in an environment free of time pressure, where the analyst can take full stock of everything that has gone into a single analysis, trying not to miss any important pieces, not letting more confirmatory information influence the decision, and assess how that evidence should weigh on the paired analytic confidence rating.

Task Complexity

Linked to the concept of time pressure is task complexity. Simple tasks and forecasts can be made increasingly difficult if time pressure is added and alternately, complex tasks can be somewhat eased if enough time is allowed to process all of the available information. These variations in complexity also, by nature, affect the amount of confidence one has when dealing with such tasks. In fact, task complexity has such a relationship with confidence that one study found it to be the *main* determinant in subjects' confidence in their responses regardless of the time limits imposed.[62] Despite

[62]Heikki Topi, "The effects of task complexity and time availability limitations on human performance in database query tasks," *International Journal of Human-Computer Studies* 62, no. 3 March 2005 [journal on-line]; available from http://ezproxy.mercyhurst.edu/login?url=http://search.ebscohost.com.ezproxy.mercyhurst.edu/login.aspx?direct=true&db=aph&AN=17516901&site=ehost-live; Internet; accessed 15 May 2007.

Heikki's finding on the relationship between complexity and confidence, no other studies were found to support the emphasis Heikki placed on it as the "main" determinant.

In a separate study from the one mentioned earlier in this literature review, Speier found that, "Decision accuracy and time were significantly correlated on simple, but not on complex tasks."[63] Though the study was more focused on the affect of interruptions during task completion, the results are still significant because when the amount of interruption was equal, complex tasks bore less accuracy.

Dunwoody et al. back up this conclusion with results from his study on decision making and human judgment.[64] The results of which again suggest task complexity hindered efforts at making accurate decisions:

> Within the confines of the current quantitative task manipulations, both task complexity and satisficing appear to explain the results. In the most complex condition, participants would have had difficulty analyzing the ecology, because its constituent components were interrelated and provided low task predictability. Conversely, the least complex condition allowed subjects to perform well with a minimal amount of cognitive effort.[65]

What Dunwoody et al. are suggesting is that people made a choice consciously or unconsciously between the accuracy of, and the amount of cognitive effort put into, their analysis. This trade-off due to task complexity carries serious ramifications for the

---

[63]Cheri Speier, Joseph S Valacich, and Iris Vessey, "The influence of task interruption on individual decision making: An information overload perspective," *Decision Sciences* spring 1999 [journal on-line]; available from http://findarticles.com/p/articles/mi_qa3713/is_199904/ai_n8837238/pg_8; Internet; accessed 15 May 2007.

[64]Philip T. Dunwoody et al., "Cognitive Adaptation and its Consequences:A Test of Cognitive Continuum Theory," *Journal of Behavioral Decision Making* 13, no. 1 Jan-Mar 2000 [journal on-line]; available from http://faculty.juniata.edu/dunwoody/jbdm.pdf; Internet; accessed 15 May 2007.

[65]Philip T. Dunwoody et al., "COGNITIVE ADAPTATION AND ITS CONSEQUENCES:A Test of Cognitive Continuum Theory," *Journal of Behavioral Decision Making* 13, no. 1 Jan-Mar 2000 [journal on-line]; available from http://faculty.juniata.edu/dunwoody/jbdm.pdf; Internet; accessed 15 May 2007.

quality of the final forecast and the appropriate amount of confidence that should be attached to it.

Heuer echoes this view on satisficing in saying:

> I would suggest, based on personal experience and discussions with analysts, that most analysis is conducted in a manner very similar to the satisficing mode. The analyst identifies what appears to be the most likely hypothesis--that is, the tentative estimate, explanation, or description of the situation that appears most accurate. Data are collected and organized according to whether they support this tentative judgment, and the hypothesis is accepted if it seems to provide a reasonable fit to the data. The careful analyst will then make a quick review of other possible hypotheses and of evidence not accounted for by the preferred judgment to ensure that he or she has not overlooked some important consideration.[66]

Heuer goes on to add that satisficing has three inherent weaknesses: focusing on a single hypothesis, failure to generate a complete set of competing hypotheses, and focusing on confirmatory evidence for hypotheses instead of disconfirming evidence.[67]

The point is that as task complexity increases so too does the potential for satisficing to occur, the results of which Heuer has outlined. The more complex the tasking is the more likely the analyst is to satisfice on some part of it, whether unconsciously or not. Though both will likely affect the ensuing confidence rating in the forecast, it is this unconscious satisficing that has the greatest potential for skewing the calibration of that confidence due to its surreptitious quality.

Source Reliability And Conflict

Source reliability, the proven accuracy of a source over the course of time, is a crucial factor to consider when developing analytic confidence. Clearly if an analyst is

---

[66]Heuer, Richards J., *Psychology of Intelligence Analysis* [book on-line].

[67]Ibid.

knowingly using consistently unreliable sources, his/her analytic confidence should be accordingly low. If the information itself is suspect, then justifying high analytic confidence would be extremely difficult.

Lee and Dry's study on decision making confidence and uncertain advice supports this assertion (See Figure 2.2). In that study Lee and Dry found that "people's confidence does not depend solely on the accuracy of the advice. Rather, confidence seems to be influenced by both the frequency and accuracy of the advice."[68]

Figure 2.2

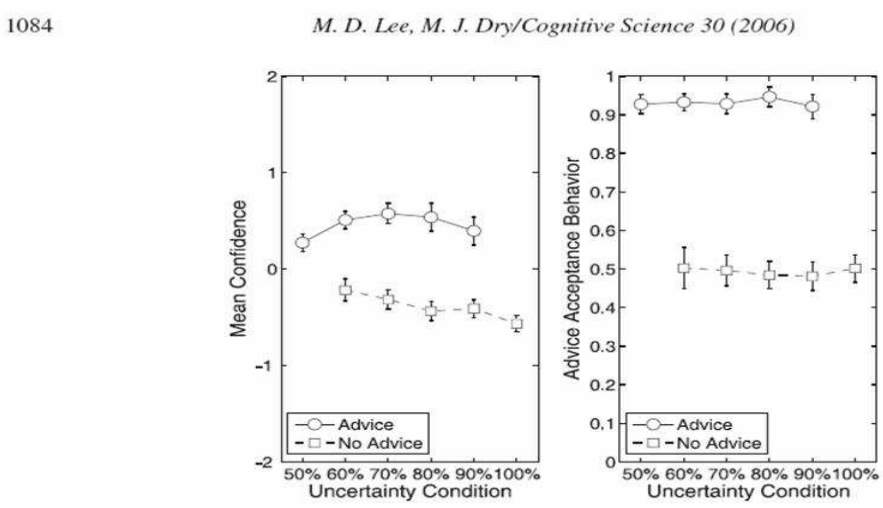


Fig. 1. The experimental data. The left panel shows the pattern of change of mean confidence for advice trials and no advice trials across the six experimental conditions. The right panel shows the pattern of change in advice acceptance behavior across the six conditions, showing the proportion of advice accepted for advice trials, and the proportion of "go left" decisions for no advice trials. Error bars show one standard error.

It is important to note that those researchers found source confidence diminished as the accuracy or the frequency of the source decreased. While this may seem surprising that subjects' confidence was not wholly placed in the accuracy of the advice, even when

[68]Lee and Dry, Decision Making and Confidence Given Uncertain Advice."

it was extremely accurate (90%), it shouldn't be. In basing their confidence on more than simple accuracy, the subjects in the experiment seem to have reserved their assessment of the advice until it proved reliable. It can be inferred that the subjects waited to see if the advice given was simply a 'lucky guess' or if it was a source in which they should place their confidence over the course of approximately 300 decisions. Based on this study it would seem that both source frequency and accuracy are needed before source reliability was developed in the eyes of the subjects.

Source corroboration, the amount to which sources conflict on a given point of information, has been shown to be another important factor of confidence calibration in forecasting. Logically, conflicting evidence, from sources of equal reliability, will likely lead to a decreased level of confidence on the part of the analyst. Emily Patterson's dissertation on analysis under data overload supports this assertion.[69] That study, which utilized 10 experienced intelligence analysts selected by their peers as being the best in their respective divisions, Patterson says that subjects used sources' corroboration and convergence to discard inaccurate pieces of evidence.[70] However, her study does make two interesting points in that corroborating evidence cannot dispel evidence from a reliable independent source, and that as time progresses the corroborative relationships between pieces of evidence and sources can change as well.[71] She also notes that analysts need to make sure the convergent information is coming from independent

[69]Emily S. Patterson, Emilie M. Roth, and David D. Woods, "A Simulation Study of Computer-Supported Inferential Analysis Under Data Overload," 12/01/99, The Ohio State University, http://csel.eng.ohio-state.edu/patterson//dissertation_tech_report.pdf (accessed May 16, 2007).

[70]Ibid.

[71]Ibid.

sources, and is not simply one source citing another.[72] A second benefit of corroboration amongst sources further boosts the advantage from having minimal evidence confliction. Not only should an analyst feel more confident about his/her forecast when the evidence isn't pointing him/her in multiple directions, but it also allows for the discarding of poor evidence that is refuted by many and perhaps more reliable sources.

Structured Methods of Analysis

Analyzing a problem utilizing a structured method has been found to improve the quality of the analysis and reduce the overconfidence that is prevalent in human intuition. Tetlock makes this point clear when saying:

> Correlates of good judgment across time and topics became more successful when the spotlight shifted from *what* experts thought to *how* they thought.[73]

This point was made in the midst of explaining that the reason why some people are naturally better forecasters than others is because of the way in which they think about and approach their analysis. In addition those forecasters who thought 'better' were also found to have better overall confidence calibration in their analyses.

It is important to clarify that this type of analysis is still based on intuition, not any type of structured approach to forecasting. The subjects of Tetlock's study were not using any type of analytic technique other than their intuitive sense of analysis.[74] Instead Tetlock's point is mentioned because it brings to light the importance of *how* the analyst approaches and conducts his/her analysis. The idea that some people are better intuitive

---

[72]Ibid.

[73]Tetlock, Philip E., *Expert Political Judgment*.

[74]Ibid.

analysts than others is of relatively little importance if everyone can improve their analysis beyond the scope of intuitive reasoning by using a structured methodology.

In his book, Heuer outlines the advantages of one particular structured method of analysis in great detail, analysis of competing hypotheses (ACH). Heuer says that one of ACH's advantages over intuitive reasoning is that it:

> Starts with a full set of alternative possibilities, rather than with a most likely alternative for which the analyst seeks confirmation. This ensures that alternative hypotheses receive equal treatment and a fair shake.[75]

The simple act of starting with a widened approach to the problem make this method, and others that offer the same advantage, a leg up on an unstructured approach to analysis. However this is not the only benefit utilizing a structured method offers. An analytic methodologies project found Heuer's ACH method "helps analysts overcome cognitive biases, limitations, mindsets, and perceptions[76]," the elimination of which should improve the final analysis and serve to better calibrate the analyst's analytic confidence. From this assertion it can be inferred that simple intuitive reasoning includes bias, limitations, poor mindsets, and perceptions, none of which would serve to better calibrate confidence.

ACH in fact addresses one such pitfall/bias specifically. Asher Koriat, of the University of Haifa, found that:

> The present studies investigated the possibility that assessment of confidence is biased by attempts to justify one's chosen answer [or

[75]Heuer, Richards J., *Psychology of Intelligence Analysis* [book on-line].

[76]Diane E. Chido, and Richard M Seward Jr., *Structured Analysis Of Competing Hypotheses* (Mercyhurst College Erie, Pennsylvania: Mercyhurst College Institute of Intelligence Studies Press, 2006).

> hypothesis]. These attempts include selectively focusing on evidence supporting the chosen answer and disregarding evidence contradicting it.[77]

Koriat goes on to say that people have a bias of favoring positive evidence over negative evidence on a given hypothesis.[78] This is where ACH truly shines in accurately calibrating analytic confidence. Not only does ACH require the listing and consideration of alternative evidence, it also scores only the *inconsistencies* with the given hypotheses, not the consistencies.

Heuer however interjects to include a word of caution in putting too much faith in the use of ACH or any other structured method:

> There is no guarantee that ACH or any other procedure will produce a correct answer. The result, after all, still depends on fallible intuitive judgment applied to incomplete and ambiguous information. Analysis of competing hypotheses does, however, guarantee an appropriate process of analysis. This procedure leads you through a rational, systematic process that avoids some common analytical pitfalls. It increases the odds of getting the right answer, and it leaves an audit trail showing the evidence used in your analysis and how this evidence was interpreted.[79]

The point being made by Heuer is that on some level all structured methods of intelligence analysis depend on the interjection of intuitive reasoning. Even the selection of which structured method to use in a given situation is a decision the analyst has to make intuitively. Though not all methods are equally robust or appropriate in all situations, the use of *any structured method* should serve to better calibrate analytic confidence by forcing the analyst to reconsider his/her assumptions and the evidence at hand. This is along the same lines as Tetlock outlined in his book when he makes the

---

[77]Asher Koriat, Sarah Lichtenstein, and Baruch Fischoff, eds., "Reasons for Confidence," *Journal of Experimental Psychology: Human Learning and Memory*, no. 6 (1980).

[78]Ibid.

[79]Heuer, Richards J., *Psychology of Intelligence Analysis* [book on-line].

point that it is more important *how* one thinks, than *what* one thinks.[80] This hint toward structure can be extrapolated into more formal methods in analysis. At the very least a structured method will move an analyst to look at the problem from a different, though maybe only slightly, perspective. Heuer also cites the elimination of satisficing as an advantage of using of structured methods.[81] He explains that ACH, and inferentially other structured methods, force analysts to consider *all* possibilities in a situation, not just the one the analyst favors or the obvious one.[82]

Structured methods also help to minimize the pitfalls intuition can have on analysis, by exposing them to the analyst during the process. Some methods may even provide an audit trail for greater oversight and accountability.[83] Furthermore, the use of multiple structured methods in an analysis, thereby further distancing the analyst from intuitive reasoning and its accompaniment of pitfalls, should increase accuracy of the analysis and better calibrate the analyst's confidence in the conclusions contained therein.

Level of Analyst Collaboration

The debate over whether better decisions, and forecasts, are made individually or in groups has been the subject of numerous studies across the fields of social sciences. Although there is still some dissent, consensus amongst the results of most experiments suggests that people tend to make better decisions, consider more possible solutions, and

---

[80]Tetlock, Philip E., *Expert Political Judgment*.

[81]Ibid.

[82]Ibid.

[83]Chido and Seward Jr., *Structured Analysis Of Competing Hypotheses*.

have better confidence calibration when working in groups as opposed to individually. Tibrewal and Poertner in their study on casework decisions in social work settings found that "overconfidence in a decision can be decreased through increasing information processing in social interactions."[84] They go on to suggest that social interaction generates greater consideration of more alternatives and overconfidence is reduced as a result.[85] This is the same reason why structured methods improve confidence calibration; it forces analysts to consider alternate perspectives in their analysis.

Scholz et al. echoes this view in a study focusing on case study judgment accuracy, where they say "studies have shown that, in a variety of tasks, group judgments are on the average more accurate than individual pre-discussion judgments."[86] The fact that these "individual pre-discussion judgments" were inferior to the group judgments means that the accuracy of individual subjects in the study improved as group discussion was incorporated into the decision-making process.

Concurring with group superiority is a study on group vs. individual decision making from Clemson University by Ahlfinger which found that "[the study's results] show strong evidence that groups working together produce more correct answers than

[84]Shradha Tibrewal, and John Poertner, "Confidence and Uncertainty in Casework Decisions: The Supervisor's Role," School of Social Work, University of Illinois at Urbana-Champaign, http://cfrcwww.social.uiuc.edu/pubs/Pdf.files/confidence.pdf (accessed May 17, 2007).

[85]Ibid.

[86]Roland W. Scholz et al., "Education in Environmental Planning: Effects of Group Discussion, Expert Information, and Case Study Participation on Judgment Accuracy," November, 2001, Swiss Federal Institute of Technology, Zurich, http://e-collection.ethbib.ethz.ch/ecol-pool/incoll/incoll_453.pdf (accessed May 18, 2007).

the same individuals working alone."[87] Then adding "[the evidence] all seems to lead to higher effectiveness of group decision making over individual decision-making."[88]

Furthermore, in perhaps the most compelling study into group decision making accuracy and confidence, James Surowiecki's book The Wisdom of Crowds is full of anecdotes of just how potent the collective accuracy and intelligence of groups is compared to single individuals.[89] From guessing the weight of an ox, to finding a lost submarine, to guessing the number of jelly beans in a jar, Surowiecki details time and again that under the right circumstances, group judgment is fair superior to even the smartest individuals.[90] These "right conditions" as outlined by Surowiecki are diversity, independence, and decentralization, all of which need to be present to harness the full wisdom of the crowds.[91] The book goes into detail describing these conditions and providing anecdotes of success or failure depending how well each instance incorporated these traits. However, it will suffice to say that all things being equal, the more people that are involved, the more diverse the group is, and the more each member acts independently, the more accurate the group's analysis will be.[92]

Despite this breadth of findings favoring group results over that of individuals, there is some evidence to suggest that at times the confidence calibration of an individual

---

[87]Ahlfinger, Hailey L., "Confidence and Work Team Performance: A Study of the Groupthink Phenomenon," 04/02/03, Clemson University, http://cujo.clemson.edu/manuscript.php?manuscript_ID=105 (accessed May 17, 2007).

[88]Ibid.

[89]James Surowiecki, *The Wisdom of Crowds* (New York: Doubleday, 2004).

[90]Ibid.

[91]Ibid.

[92]Ibid.

may be superior. In a study on confidence in individual and group decision making, Puncochar and Fox found that:

> Under conditions of group work and instructor feedback, students produced higher exam accuracy scores working in groups than alone but at a cost of increased confidence for groups' wrong answers. Groups' high confidence for wrong answers generated the case when "two heads' are worse than one." Students participating in groups that arrived at wrong exam answers gave higher confidence when wrong and lower confidence when correct for repeated items on a final exam.[93]

Though these findings directly conflict with the previously cited studies, it should be noted that Puncochar and Fox used multiple choice tests and subjects first took them individually before grouping up to take the same test again. As a result it is possible that individuals with high confidence individually were able to persuade fellow group members to even further increase the group's confidence in right or wrong answers. The logic of collective intelligence appears to have failed in this case. Puncochar and Fox offer another reason why group confidence may have been detrimentally high:

> Multiple-choice items offered four choices, usually the groups focused on two options, which suggests that groups would be likely to indicate more certainty.[94]

It should be noted that the tests had 4 answers to pick from, one of which being completely right while the others were completely wrong, deviates greatly from intelligence analysis and its perpetual ambiguity.

Much of the debate over the optimal level of collaboration in accurate decision making, and associated confidence ratings, focuses to varying extents on the concept of groupthink. The term was coined by psychologist Irving Janis, who spent a great deal of

---

[93]Judith M. Puncochar, and Paul W. Fox, "Confidence in Individual and Group Decision Making: When "Two Heads" Are Worse Than One," *Journal of Educational Psychology* 96, no. 3 2004 [journal on-line]; available from http://www.psychologia.pl/~jasia/puncochar.pdf; Internet; accessed 18 May 2007.

[94]Ibid.

time researching group decision making. Groupthink is a phenomenon that occurs "when a group makes faulty decisions because group pressures lead to a deterioration of mental efficiency, reality testing and moral judgment."[95] Janis goes on to add:

> Groups affected by groupthink ignore alternatives and tend to take irrational actions that dehumanize other groups. A group is especially vulnerable to groupthink when its members are similar in background, when the group is insulated from outside opinions, and when there are no clear rules for decision making.[96]

Clearly this concept poses a large danger to group decision making and analyst collaboration in intelligence analysis. Examples succumbing to groupthink abound in world history and intelligence failures. The failure to anticipate an attack on Pearl Harbor, the unintended outcome of the Bay of Pigs invasion, and even the claim of Iraq possessing weapons of mass destruction are all examples of how groupthink poisoned collaborative decision making and the confidence placed therein.[97]

However, over the course of his career Janis was able to document some symptoms of potential groupthink within groups, as well as remedies for it once it has manifested itself.[98] These findings, as seen in the table, logically lead not only to a more dynamic and functional group, they also read as a list of recommendations by which intelligence analysis could be improved when done in groups (See Figure 2.3). In fact, becoming familiar with the symptoms and countermeasures may also lead to better, more objective intelligence analysis.

---

[95]Irving Janis, "What is Groupthink," Psychologists for Social Responsibility, http://www.psysr.org/groupthink%20overview.htm (accessed January 7, 2000).

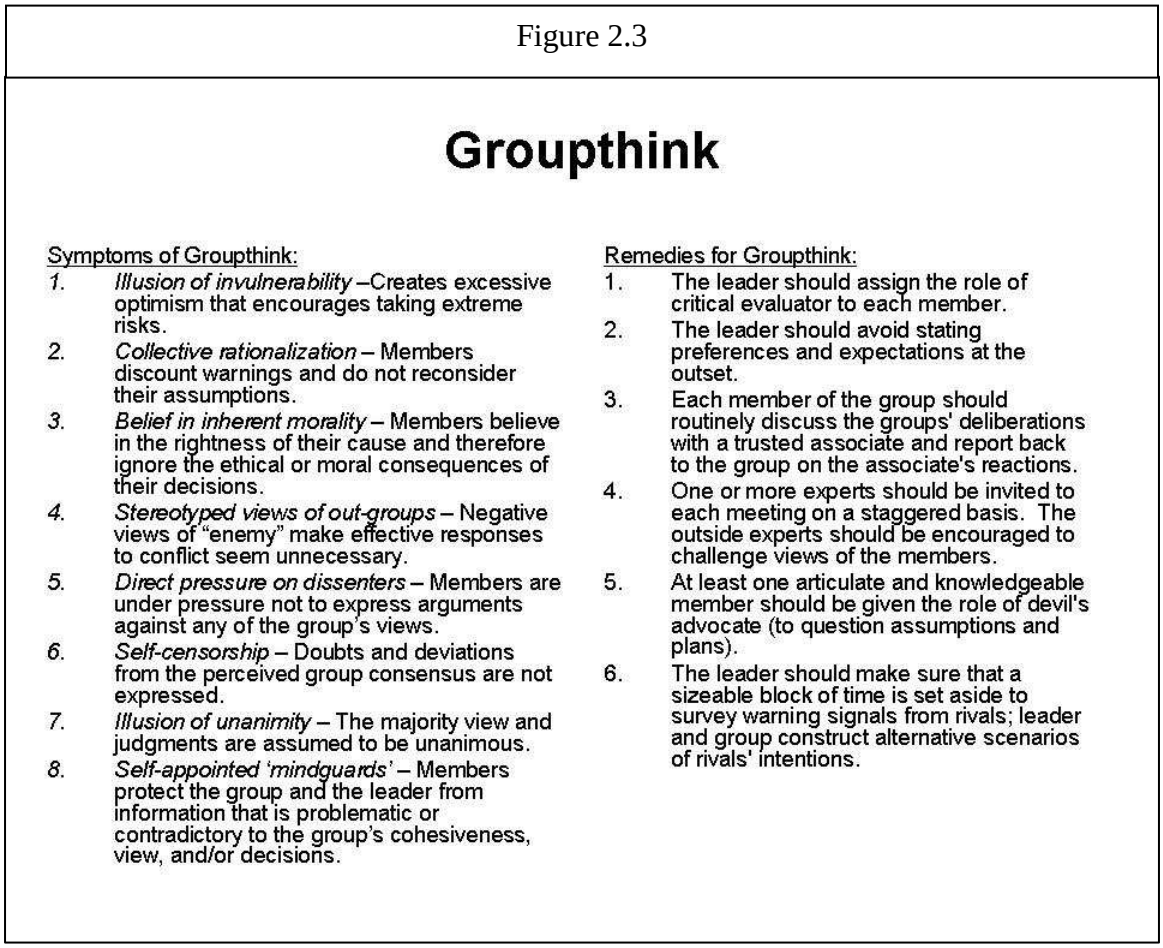
[96]Ibid.

[97]Ibid.

[98]Ibid.

The implications of adhering to Janis' findings could include not only better calibrated confidence ratings, but also the creation of a set of goals for which analytic groups in the intelligence should strive to meet. By being cognizant for the signs of groupthink, and being well versed in how to minimize its effects, a high level of collaboration becomes an appropriate factor to measure when constructing an analytic confidence rating.

Figure 2.3



Based upon the factors previously discussed and their correlation with increased confidence and confidence calibration, this researcher felt that these factors were

appropriate factors to consider when formulating a rating of analytic confidence in intelligence analysis.

CHAPTER 3:

METHODOLOGY

In order to test the given hypotheses I had to conduct an experiment. The experiment tested whether the factors of constructing a properly calibrated analytic confidence rating as outlined in the literature review are indeed appropriate and valid measures. This experiment was designed to determine if intelligence analysts were able to more accurately rate their analytic confidence when the factors identified in the previous chapter were present, compared to when they were not. This methodology section will detail the research design of this experiment.

Research Design

The experiment I conducted broke my subjects into three groups (see the next section for a full description of the subjects). Two of the groups were experimental groups, and one was a control group, used to set a baseline with which to measure the other two against. The experimental groups were subjected to differing conditions designed to control for the combined independent variables as designated in Chapter 2 (time, group work, source reliability, structured methods etc). The Low Confidence group was given 15 minutes for their analysis, worked alone, had low source reliabilities, and could not use structured methods. The High Confidence group worked alone and in teams, used a structured method, had high source reliabilities, no time pressure, and was encouraged to discuss the scenario with others in the group.

Subjects

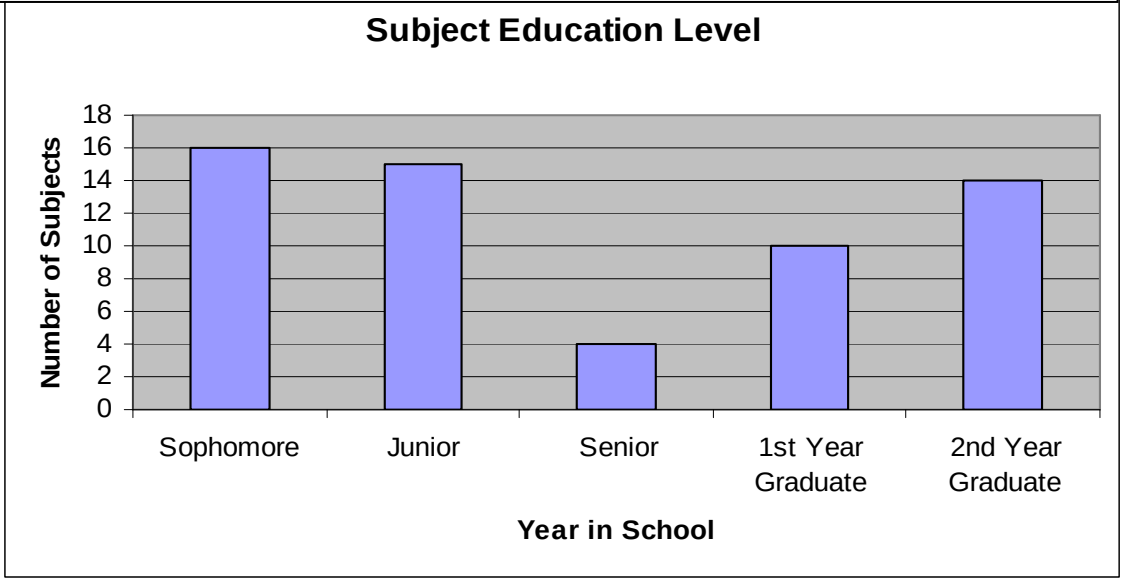
In order to better apply the findings of my experiment to the United States Intelligence Community, I chose undergraduate and graduate students at the Mercyhurst College Institute for Intelligence Studies. This program prepares students for careers in the intelligence analysis field. Due to security clearances and classification problems, it would be very difficult to utilize actual analysts within the intelligence community, so these students were viewed as an appropriate alternative sample group.

Mercyhurst College is the oldest and largest institution in the world which offers a four-year program in Intelligence Studies, with a focus toward producing entry level analysts upon graduation. Students in the program are subjected to a rigorous academic curriculum from their first semester enrolled through graduation. The program prepares students for entry-level intelligence analysis in the fields of national security, law enforcement, and competitive business intelligence. In addition to foreign language proficiency requirements, numerous projects, fast paced classes, and internships, students are challenged to work on real world intelligence projects utilizing open source intelligence. Thus seniors within the program are competent, capable analysts possessing a wide variety of analytical skills and abilities. This analytic education, skill set, and experience makes them an ideal population on which to test my hypothesis without security clearance impediments.

To get participants for my experiment I sent out an E-mail to all of the students within the Intelligence Studies program at Mercyhurst College, both undergraduate and graduate students and posted flyers in the Intelligence Studies building (See Appendices 1 & 2). The E-mail contained information about the time, place, and topic of the

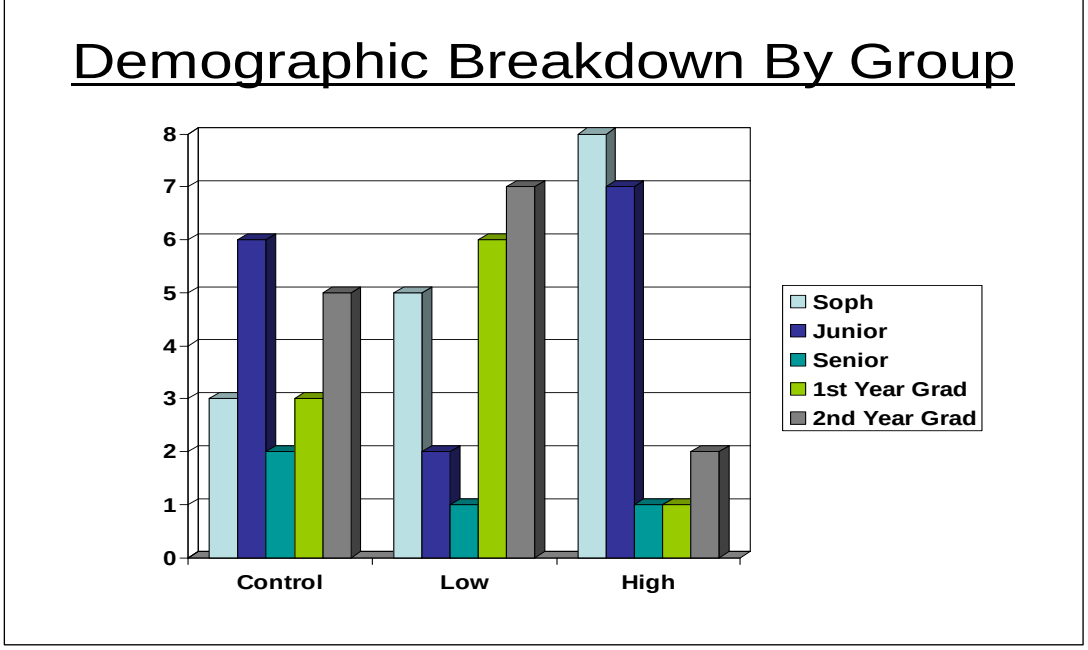
experiment, in addition to asking students to sign up on a bulletin board near the classrooms in the Intelligence Studies building on campus. The E-mail stated participation was welcome by anyone other than freshman students, and that most professors within the Intelligence Studies Department would be offering extra credit in their courses to those students who participated. Freshman students were left out of the experiment due to their not having learned the Structured Analysis of Competing Hypotheses method which was to be used in one of the experimental groups. The exclusion of freshman students also ensured a slightly more experienced and mature cadre of participants. The sign up sheets for the experiment were set up in 3 1-hour blocks, with 21 slots per block (See Appendix 3). Subjects were free to sign up for any block they wished, which was done in an attempt to ensure a random distribution of demographics between the 3 groups. Testing took place 2 weeks into the winter term, during which time students have more free time owing to a lack of tests that early into the

Figure 3.1



term. Although all students should have been on relatively equal academic footing that early into the term, it is possible that the extra credit was enough of an incentive to entice some students to participate who otherwise would not have. A total of 60 students participated in the experiment, with a fairly equal showing across the 5 strata of educational levels, save a noticeably smaller number of seniors. A breakdown of the subjects by educational level is found below (See Figures 3.1 & 3.2).

Figure 3.2



Process

Before conducting the experiment I had to get consent from both the Mercyhurst College Institutional Review Board (IRB), and the participants themselves. Any researcher conducting an experiment involving human subjects at Mercyhurst College must be granted permission to do so by the IRB, which requires researchers to outline

their experiment and identify any potential dangers or hazards which participants may come into contact with as result of participating in the experiment (See Appendix 4). Upon IRB approval I next had to secure the subjects' consent to participate in the experiment. Once all of the students in each group had settled into their seats, I passed out the Consent Form (See Appendix 5). The form briefly outlined what the experiment entailed, that there were no foreseeable dangers to the subjects, and that they were free to leave at any time. The Consent Form also asked for some basic demographic information, as well as the names of the subjects' professors, in order for me to pass along the names to the appropriate professors so the students could be awarded any applicable extra credit.

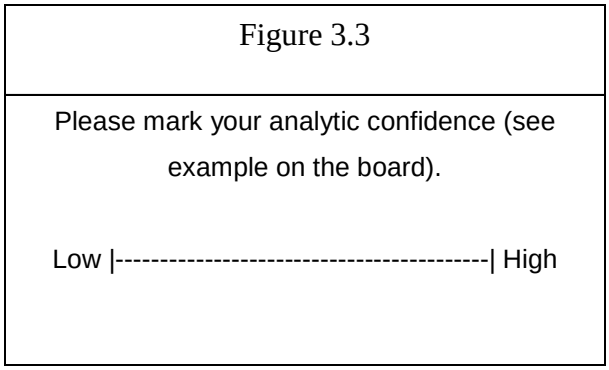
Before starting each group's section of the experiment, I asked if there were any questions, and instructed each section to carefully read the instructions on the sheets handed to them.

Control Group

The first group to participate in the experiment was designated to be the control group. This group was given the same fictional intelligence scenario (See Appendix 6) as the rest of the groups, and was instructed to simply write an analysis of the scenario. Each scenario, regardless of group, contained a brief set of instructions, the scenario and evidence, room to write an analysis, and the continuum to rate analytic confidence. However, the experimental groups were also given definitions of key terms such as analytic confidence and source reliability in order to help them understand what was expected of them. On the control group's sheet, there were no ratings of source

reliability in order to leave everything as manipulation-free as possible. This group was given as much time as they felt they needed, though it is likely they felt a very slight time pressure as they had only signed up to participate for approximately 1 hour's time. After reading the scenario, writing their analyses, and rating their analytic confidence by way of a slash through a continuum-like scale (See Figure 3.3), the subjects' participation obligation was completed. The use of a continuum-like scale to rate analytic confidence is unique, and was utilized in order to avoid the foreseeable problem of most students automatically/unconsciously rating their analytic confidence in the range of 5-8, as many professors told me I should expect. This norm amongst students of rating their analytic confidence within approximately 5-8 range is the result of students' realization that professors will question them on their confidence if they score it any higher than roughly an 8. Concordantly, students are also wary to list their confidence as less than a 5, as doing so would put them on the firing line with professors questioning them about how useful it would be to a decisionmaker and why they are taking time to present an analysis they gave a confidence score of 2 or 3 to. Subsequently, I thought the continuum method to be a better way of soliciting thought and more honest self-evaluations of analytic confidence due to its uniqueness, than merely asking for a number on a 1-10 scale as is typically done at Mercyhurst College. Looking at the results of the experiment, I feel the continuum method of scoring proved to be very successful in eliciting scores across the entire 1-10 scale. The subjects then

Figure 3.3



read a Debriefing Sheet (See Appendix 7), thanking them for their participation, passing along contact information for future reference, and instructing them not to discuss what they had just done with anyone else in order to keep the rest of the groups' subjects as unbiased as possible. The subjects were then free to leave.

Experimental Group 1: Low Confidence

The second group to participate in the experiment, and the first experimental group, was designated as the Low Confidence group. In this group, though the exact same scenario was handed to the subjects, source reliabilities were rated low, a stern 15 minute time limit was imposed on the subjects, and they were told not to speak to one another while performing their analysis. In addition, to increase the amount of confliction between sources, evidence that went against the majority of the facts was given slightly higher source reliability. The process by which the group went through the experiment was the exact same as the control group, with the notable exception being the time limit imposed. The independent variables were manipulated to be absent, as they had been identified by literature (see Chapter 2) as being valid factors in increasing confidence.

Experimental Group 2: High Confidence

The third group to participate in the experiment, and the second of the two experimental groups, was designated as the High Confidence group. In this group, the exact same scenario was handed to the subjects, however source reliabilities were rated much higher. Additionally, no time limit was imposed on the subjects, they were

instructed to use the Analysis of Competing Hypotheses software found on the computers (with which all subjects were familiar due to having used it previously in class), and they were instructed to work in groups in order to discuss their thoughts/analysis of the scenario (See Appendix 8 for a detailed discussion on ACH). To break the subjects in this group up into smaller subgroups to discuss their analyses, I numbered the subjects off and split them randomly into groups of 4's and 5's. Students did not know I would do this so it is unlikely any of them attempted to manipulate the group to which they were assigned. This level of group work allowed the subjects to come up with their own analysis, discuss/collaborate with other subjects, and then go back and re-craft anything they wished to change. The independent variables were manipulated to be present, as they had been identified by literature (See Chapter 2) as being valid factors in increasing confidence accurately. Despite the differences in the independent variables, the experiment was conducted in the same manner as described above in the Control Group section, starting with the Consent Form and ending with the Debriefing Form.

Data Analysis Procedures

Since the intent of the experiment was to test whether or not the factors identified in the literature review were appropriate factors in assessing analytic confidence; ratings of analytic confidence were measured on a continuum-like scale 5cm long. Students were asked to draw a line or slash over spot on the continuum line which indicated their amount of analytic confidence. Then, to convert the slashes to a numeric score, I took a ruler and measured where the subjects' slashes crossed the continuum and recorded them. For example: If a slash crossed the continuum 3.5 centimeters from the left side (lower

confidence) of the continuum's starting line, then that subject's analytic confidence was recorded as a 7. The measurement was multiplied by 2 because the continuum was only 5cm wide, so each conversion was the measurement from the continuum's leftmost point, and then multiplied by 2.

After measuring all of the subjects' analytic confidence scores, I put them into a Microsoft Excel Spreadsheet along with each group's demographic data, and began my statistical analysis to see if the experiment's results were statistically significant.

Amongst all three groups the analytic confidence ratings expected varied greatly. The control group was expected to be all over the continuum in their ratings, while the low and high experimental groups were expected to be near their designated end of the continuum.

CHAPTER 4:

RESULTS

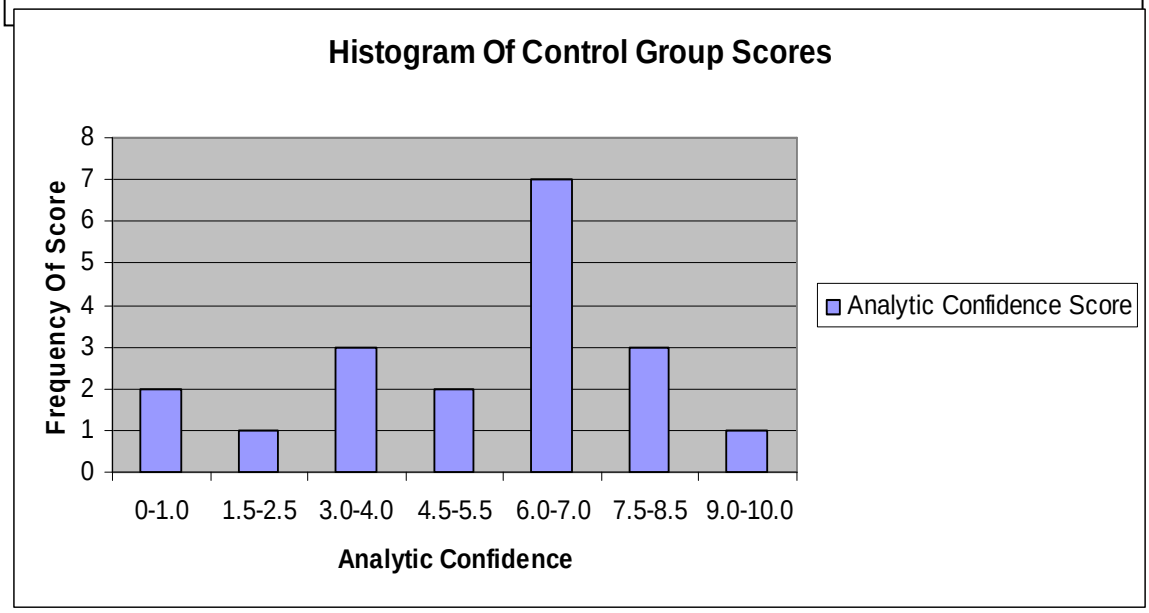
The results of the experiment on analytic confidence factors yielded a number of interesting and a few somewhat surprising results. This chapter will detail the results of the experiment, both relating to each group individually, and compared to each other as a whole. For all three groups, statistical analysis was performed to determine if the findings uncovered were significant, and in one case they were.

Control Group Findings

As discussed in the previous chapter, this group was included to provide a baseline against which to compare the other two experimental groups. It also served as a way to see if subjects who presumably understood very little about analytic confidence would uniformly rate the same scenario equally or be all across the continuum in their ratings. The latter proved to be the case. The scores of the 19 subjects in the Control Group ranged from 1.0 to 9.0 with a mean score of 5.4. (See Figure 4.1) This group was also the most equally distributed in terms of demographics (See Figure 3.2) with no one or two years commanding an overwhelming majority. This group's results were what I had predicted would be the case: That all things being equal, analysts' analytic confidence will be all over the board (See Appendix 9). However, there is one interesting result to come out of the control group, an average of only 5.4. In spite of abundant literature suggesting that people are naturally over confident, and my own expectation of a slightly higher mean (6-7.5 due to students' typical analytic confidence ratings on assignments), the mean was a modest 5.4. (See Figure 3.2 for group demographics)

Although the possibility that the continuum method of rating confidence was the cause of this remains an open question, I do not think it would lower the mean an entire point or more on a 10 point scale.

Figure 4.1

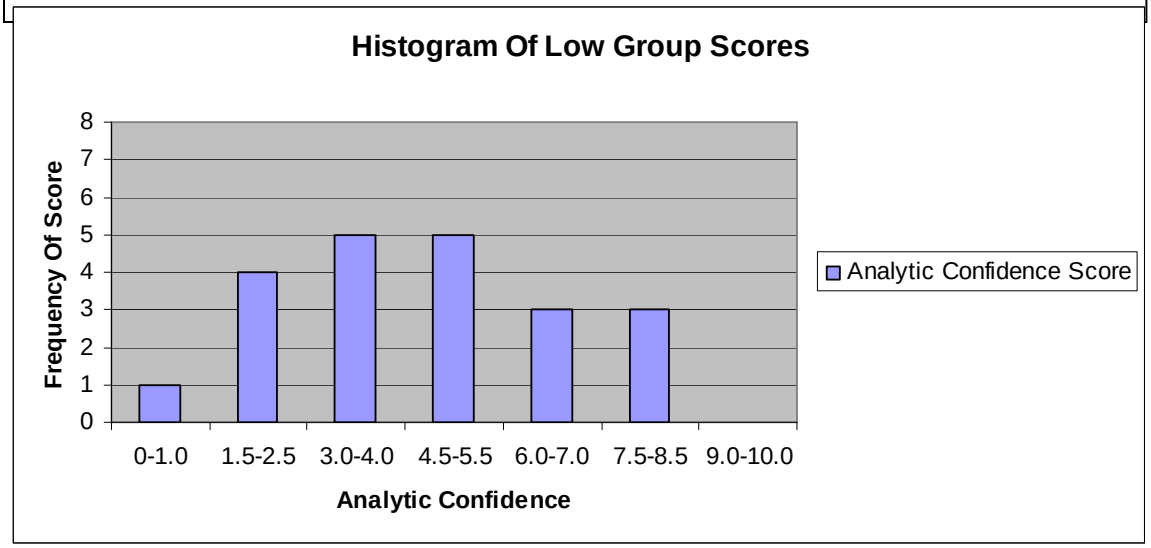


Experimental Group 1: Low Confidence Findings

As discussed in the previous chapter, this group's independent variables were manipulated so as to significantly decrease the appropriate level of analytic confidence based on the literature review in Chapter 2. This group was expected to have significantly lower rating of analytic confidence compared to the Control Group, but more importantly the High Confidence Group. The scores of the 21 subjects in the Low Confidence Group ranged from 0.5 to 7.5 with a mean of 4.4 (See Figure 4.2). Despite 3 scores of 7.5, the results were generally as expected, with a lower mean and an array of

scores near the lower end of the continuum. The difference between the control group mean and the low confidence mean was not statistically significant. However, the narrower range and lower high score indicate that variables affecting analytic confidence, which were absent for this group, did have an effect on the analyst's analytic confidence. Demographically, the majority of this group was graduate students (See Figure 3.2), though their scores shared the same range as the entire groups' suggesting this plethora of postgraduates did not affect the group's average.

Figure 4.2

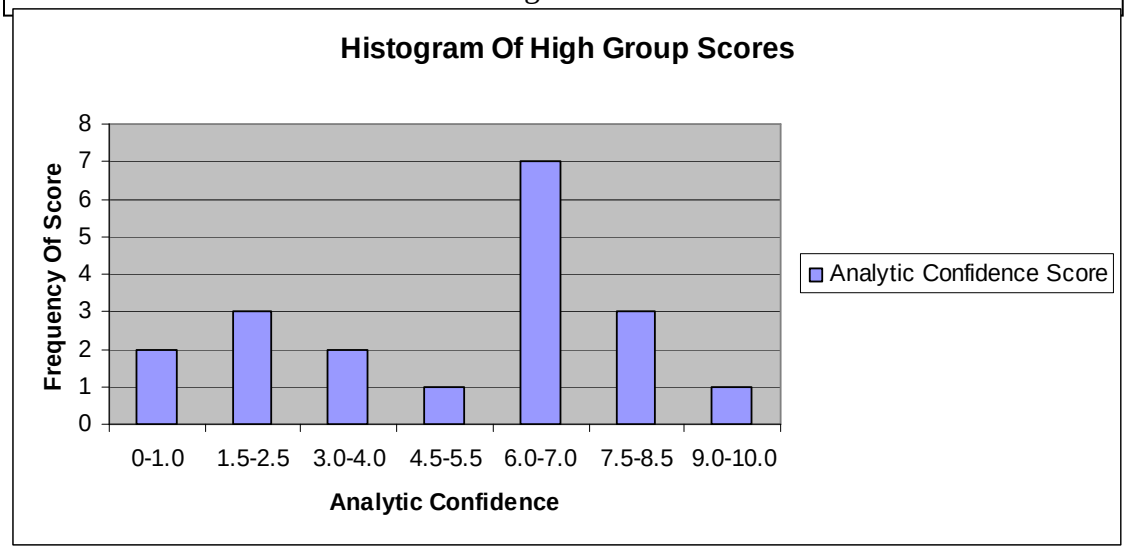


Experimental Group 2: High Confidence

As discussed in the previous chapter, this group's independent variables were manipulated so as to significantly increase the appropriate level of analytic confidence based on the literature review in Chapter 2. This group was expected to have significantly higher analytic confidence compared to both the Control Group and the Low Confidence Group. The scores of the 19 subjects in the High Confidence Group ranged

from 1.0 to 9.5 with a mean of 5.3 (See Figure 4.3). These numbers bear scores of important results to discuss.

Figure 4.3



First the range, which is wider than even the Control Group's, is surprising indeed. Despite many of the variables found to increase (and help to better calibrate) analytic confidence being present, 5 of the 19 students gave analytic confidence scores of 1.0, 2.0, or 2.5, of which 2 were juniors, 2 were sophomores, and 1 was a second year graduate student! Interestingly, the analysis these subjects wrote wholly lent itself toward a very high level of analytic confidence. The analyses were well laid out, used Words of Estimative Probability (WEPs) superbly, and identified a few of the weak points in the evidence given, yet extremely low scores were given by all 4. While the subjects who gave their analytic confidence as 2 or lower were not statistical outliers, they were excluded during the statistical analysis of the experiment due to their analysis and analytic confidence being wholly contradictory. Post experiment interviews were conducted with these subjects, and none could articulate what in the information given in

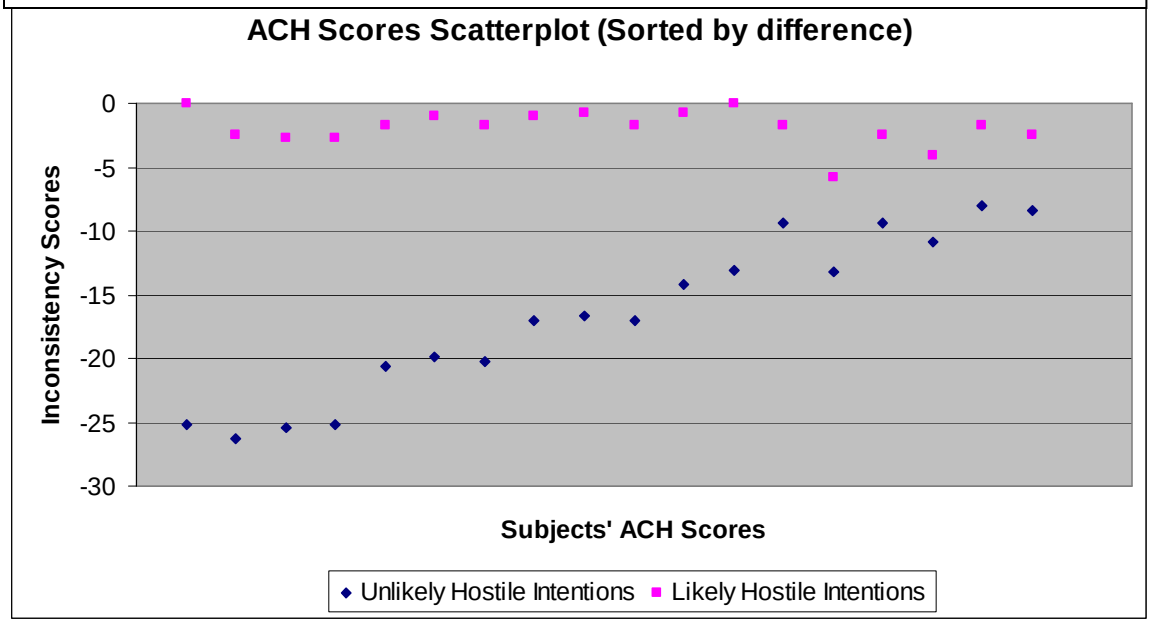
the experiment led them to such low scores. The only common thread amongst their responses was that they felt that despite the overwhelming amount of evidence, high source reliability, and group work consensus, the evidence was circumstantial at its core. Even if one accepts this view, to actually present such an analysis based on the evidence in the experiment, to a real world decisionmaker would be precarious at best. In reality, none of these subjects could reasonably justify the difference between their analytic confidence and their strongly worded analyses, so those results were excluded.

Second, the demographic split of this group initially suggested that the lower average could be attributed to the youth of the subjects (See Appendix 9). With 15 of the 19 subjects being sophomores and juniors, it is possible that more experienced student analysts would have rated their analytic confidence higher on as a whole. However the results of the older subjects do not show this speculation to be valid. Moreover, in none of the 3 groups did years in school seem to have any drastic effect, with each year having a wide dispersal in scores.

Third, the scores of the Analyses of Competing Hypotheses (ACH) software is extremely interesting. ACH is an analytic methodology in which hypotheses compete against each other as evidence that is inconsistent with one hypothesis or the other is tallied. In a nutshell, the methodology centers the analyst's attention on evidence which is inconsistent with a particular hypothesis, and the more evidence that is inconsistent with a hypothesis, the less likely it is to be right. To begin with, the scores of the students' ACH's were not commensurate with the level of analytic confidence rated, with the exception of a lone student whose analytic confidence was rated at 9.5. The rest of the group had such scores at -2.414 to -26.312 and -0.707 to -16.656, yet rated analytic

confidence at 6.5 and 1 respectively (See Figure 4.4). Even the most lopsided score, 0 to -25.191, garners only an analytic confidence score of 6 of the subject.

Figure 4.4



ACH itself is a structured analytic method proven to improve analysis, and the nuance therein. While there is no 'typical' score for an ACH matrix, this author has found that with most analyses there usually is no more than 7-8 points between the two hypotheses being tested. In fact, an analysis which produces a spread of more than 5-6 points can begin to border obvious, and thus it may need to be refocused and narrowed even further. With this in mind, this author's confusion surrounding the ACH scores and their paired analytic confidence ratings comes to light. The ratings of analytic confidence in relation to these impressively one-sided SACH scores is confusing at best, though perhaps illogical is a better label. Perhaps the best question to ask of this result, and regrettably one this researcher did not have the opportunity to ask is: If an ACH score of 0 to -25.191

only makes your confidence a 6, what score would it take to garner a 9.5? Clearly there is some confusion, lack of understanding, or simple carelessness present.

Summary of Results

The data from the experiment was analyzed for statistically significant results using a program called Statistical Package for the Social Sciences (SPSS), with the assistance of a statistics professor at Mercyhurst College. The output of the program was a large and complex discussion of numerous statistical tests and formulas, however only one of these is pertinent for statistical significance in the context of this experiment (See Appendix 9).

Results of the SPSS testing on the analytic confidence scores across the three groups indicate that there is a statistically significant difference between the means of the high and low confidence group at the 95% confidence level. The significance score in this case was .038, which would actually put the confidence level at 96.2% to be more precise. (See Figure 4.5) What this significance means is that there is approximately a 3.8% chance that these results were sheer coincidence and not caused by the variables in the experiment. While the difference in means is fairly small, < 2 points on the analytic confidence scale, a confidence interval of 95% indicates that this should not be attributed to chance, and instead is due to a statistically significant difference in the data from the experiment.

Figure 4.5

SPSS Statistical Output of Experimental Groups
Statistical Significance is Outlined at the Bottom

Multiple Comparisons

Dependent Variable: Analytic Confidence
Scheffe

(I) Section	(J) Section	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
High	Low	1.92857*	.73046	.038	.0878	3.7693
	Control	.91228	.74631	.479	-.9684	2.7930
Low	High	-1.92857*	.73046	.038	-3.7693	-.0878
	Control	-1.01629	.68414	.339	-2.7403	.7077
Control	High	-.91228	.74631	.479	-2.7930	.9684
	Low	1.01629	.68414	.339	-.7077	2.7403

*. The mean difference is significant at the .05 level.

Output table from SPSS statistical analysis, the level of significance in the "Sig" column shows there to be a statistically significant results at the 95% confidence level.

The findings discussed in the previous sections indicate that students, and by some implication analysts universally, do not understand just what analytic confidence is, nor what are valid factors to consider when developing an accurate rating of it. Even in an experiment which subjects were given the same information (though the context in which it was presented was different), their confidence scores varied greatly within all three groups. This could be the result of a simple lack of understanding as to what analytic confidence really is, and how they should assess it. This lack of understanding, on some level, is the result of a dearth of instruction on the topic by professors at Mercyhurst College. This, in part, was the impetus behind this work. There is some more promising results to take from this study however. The results of the high and low

confidence groups indicate that without even realizing it, students have a vague notion of what should and shouldn't influence their analytic confidence. They may not be able to articulate all of the factors they unconsciously considered, but nevertheless the spread between the two groups is evidence enough. With better instruction at Mercyhurst College, students may be able to refine their analytic confidence ratings, given this and future research on the subject. These same results also highlight the need for a more methodical approach for developing analytic confidence ratings in intelligence estimates. The results show that intuitively, analysts vary greatly across the spectrum of scores, including very low scores when their confidence should be quite high, and reasonably high scores when their confidence should be much lower.

There are some methodological changes which could be made to make the experiment more valid. The final chapter will discuss some suggestions for improvement of the experiment, the implications of the experiment's results, and highlight areas for continued research in the future.

CHAPTER 5:

CONCLUSIONS

The purpose of this study was to determine appropriate factors to consider when developing a well calibrated analytic confidence score. By controlling for variables discussed in Chapter 2 which were found to validly increase or decrease confidence in decision making, subjects were split into groups where many or few of these variables were present. While the Low Confidence group's mean confidence was a modest 4.4, and the Control group's mean confidence was a scattered 5.4, the High Confidence group did not produce a proportionately high mean response. Though the results of the High Confidence group are perplexing, especially given the associated ACH scores discussed in the previous chapter, the independent variables (i.e. the factors discussed in Chapter 2) manipulated in the experiment *cannot be ruled out* as appropriate factors in formulating analytic confidence. The other two groups in the experiment produced scores which strengthen the argument that factors such as time, source reliability, team work, and source corroboration are indeed appropriate things to assess when rating one's analytic confidence, as they had the expected affect on the analysts' analytic confidence. Most notably, the Control group's wide range of scores indicates the confusion amongst analysts in just what an appropriate analytic confidence score is, and the pitfall of relying wholly on intuition known to be faulty. Examining only the control group, it is troubling to put oneself in the position of a decisionmaker being given such a wide range of analytic confidence from the same information. Clearly a better and more accurate method of rating analytic confidence needs to created, if only to narrow the spectrum of scores given the same information.

One method of addressing the confusion surrounding the scoring of analytic confidence would be to include a numerical expression of analytic confidence in each finished intelligence product. This numerical expression would express the analytic confidence an analyst has in his/her forecast. This figure would then be included at the end of each analysis and could be set on a 1 to 10 scale, as was done by way of measurement in the experiment's results discussed in the previous chapter. 1 being extremely low confidence based on the appropriate factors described previously, and 10 being the infamous "slam dunk," though again based on the factors identified and proven to appropriately affect analytic confidence. In this way, readers understand the analyst's level of confidence clearly and succinctly and analysts would have a uniform method of rating their confidence. Rating methods similar to the one suggested above have circulated throughout both the intelligence community and academia, as mentioned previously in Chapter 2 of this thesis.[99]

A second conclusion to extract from this study is that subjects were more willing to commit themselves to lower scores than higher scores. Whether they felt more secure in lower confidence, as it would be easier to explain if questioned, or merely hesitant to go out on an analytical limb and mark high confidence in an experiment they knew was designed to measure analytic confidence, remains unknown. It is possible however, that some of these students were aware of the literature regarding overconfidence, and thus did not want to appear overconfident in an experiment focusing on the same topic.

Another important conclusion to draw from this study is the apparent lack of understanding of the relation of ACH (and maybe structured methods in general) to

[99]Michael Schrage, "What Percent Is 'Slam Dunk'?," 02/30/05, Washington Post, http://www.washingtonpost.com/wp-dyn/articles/A37115-2005Feb19.html (accessed May 14, 2007).

confidence on the part of the subjects in the High Confidence group. At least one subject's ACH matrix demonstrated a clear misunderstanding of the fundamentals of ACH, however the correlation of other subjects' analytic confidence to their ACH scores indicates the problem was more widespread. Even professors within the Intelligence Studies Department at Mercyhurst College, who teach ACH to analysts outside the school, were at a loss to explain the extremely low scores given by subjects with such one sided ACH scores. That students would score their analytic confidence so low in the experiment seems nearly hypocritical based on this researcher's knowledge of students' past analytic confidence scores dealing with much more difficult subjects and more conflicting evidence than presented in the experiment. The one possible explanation for this phenomenon is that the subjects in the experiment felt as though there was relevant evidence/information left out of the scenario handout that they should take into consideration. That is to say that the High Confidence group thought the scenario given was too one sided and that they should be cautious when rating their analytic confidence in order to avoid appearing overconfident. A result of which could be the subjects gaming the experiment, and not accepting instructions that the only information available for their analysis was given to them. While this goes against the instructions given to the subjects, this does make some sense given that analysts are naturally prone to seeking out as much information as feasible on a given topic.

There also seems to be some level of intuitive understanding of analytic confidence amongst students in the intelligence program at Mercyhurst College, and likely the US intelligence community as a whole. It is sort of a 'I know it when I see it' level of understanding, not very specific or deep but the experiment's results reflect a bit

more than a minimum grasp of this concept. With more instruction and continued research on this topic, rating analytic confidence could become second nature to analysts.

A final conclusion to draw from this study is that there is a growing interest amongst intelligence studies students, and by implication future analysts, to more fully comprehend analytic confidence. The reasonably large turnout for the experiment, on short notice just after a break in the school year, demonstrates that the relative dearth of information provided to them on analytic confidence in the classroom has aroused some amount of curiosity on the subject. While it is true that the extra credit offered by some professors for participating may also be at play, it is important to note that 2 of the subjects came knowing they would receive no credit for their time. It is also noteworthy that past experiments conducted by graduate students at Mercyhurst College have not consistently gotten this many subjects to sign up and participate.

As mentioned in the previous chapter a couple of changes in the methodology of the experiment might produce more reliable and significant results. The following changes to the methodology might yield more robust results in support of the hypothesis. To begin with, a larger number of subjects would be preferable. As seen in the results of this study, a few scores at either end of the spectrum weigh heavily on the mean and statistical significance of the group's overall score. Furthermore, a larger number of subjects would include more undergraduate upperclassmen, especially seniors, of whom there were few in the experiment. Although the analytic confidence scores of the subjects in all 3 groups of the experiment did not correlate with subjects' year in school, a more diverse and random sampling would be ideal. Clearly a sampling of professional

analysts from within the US Intelligence Community would be ideal, but that was outside the realm of possibility for this study.

Additionally, a refresher presentation on the use, utility, and robustness of ACH prior to passing out the scenario to the High Confidence group may yield analytic confidence score more commensurate with the ACH scores found in this experiment. Additionally, going over the key term definitions with both experimental groups may help future subjects to more carefully consider their analytic confidence scores, as this would reduce the likelihood that subjects merely skimmed over that section of the handout.

Finally, the use of a continuum line versus a simple numeric expression for the subjects' analytic confidence could potentially yield different results. True, subjects should have been able to mentally estimate where the halfway point on the continuum was and then mark their scores accordingly, but sticking with the same numeric expression the subjects were used to might change some of the scores. While it is only minimally more difficult to write a number, a simple slash on a continuum can be done carelessly and without thought, so perhaps asking for a numeric expression would be more appropriate in a second experiment.

There is certainly a need to continue research on analytic confidence, especially to conduct this experiment again, ideally with a more ideal population of subjects, in order to more fully test the hypothesis found in this thesis. Additionally, one specific question to be researched is: what is the optimum level of knowledge or experience on a given topic that will yield the best results and confidence calibration? There is also a need to examine specific factors discussed in previous chapters such as source reliability. How

should one come up with a numeric expression of it? What should that score be based upon? What causes that reliability to change and when it does, how much should it change? Questions like these are crucial to a full understanding of just how analytic confidence should be assessed, for without a deeper understanding of its parts, analytic confidence as a whole may well continue to be confusing to analysts and decisionmakers alike. Additionally, further research is needed by those social scientists with access to the pool of analysts within the US Intelligence Community, as results from that population would more applicable to intelligence analysis. Heuer certainly got the ball rolling on this, but recently there has been very little work done at the Center for the Study of Intelligence on analytic confidence that has been made public.

A Step Forward

Throughout this thesis there have been numerous mentions of the need for a more structured method by which to calculate a score of analytic confidence in intelligence estimates. After having identified and discussed some of the appropriate factors to consider when developing analytic confidence ratings, and conducted an experiment which at least partially affirms the hypothesis, this research set about trying to develop just such a method. Taking into consideration the pitfalls of relying on human intuition, and the confusion of such wide ranging scores of analytic confidence in the Control group, this researcher created a self assessment table for developing analytic confidence systematically (See Appendix 10).

The table is essentially list of the factors identified in the literature review as having been found to valid measures of confidence in decision making, an associated

numeric range on which to score oneself, and a box to add up the scores from all of the factors. In order to derive an analytic confidence score, an analyst simply rates herself on each of the factors individually, within the range indicated, totals all the scores, divides by 45, the total number of 'points' possible, and multiplies by 10, in order to produce a number on a 1-10 scale.

While there are certainly a number of foreseeable flaws or pitfalls to this method, it at least appears to be a step in the right direction toward narrowing the range of analytic confidence scores amongst analysts, and focusing them on appropriate things to consider when rating their own analytic confidence. It will not force one score on all analysts, as it is still very subjective at its core, but it may tend to keep analysts relying solely on how confidence they think they should be. Additionally, the explanations provided next to each score should reduce confusion associated with each sub-score amongst analysts. An even greater benefit of these explanations could come by way of attaching this self score sheet to the report it is based on when handing it to a decisionmaker. Not only would it provide an analytic confidence rating, it would also immediately show how that score was arrived at (in an easy to understand table), and it would be a tangible record for all involved to examine in the future.

Admittedly, there are a number of potential problems with this method. First, it still relies on human intuition and objective, impartial self-scoring, which an analyst could 'tinker' with if he did not like the initial overall score that he came up with. Second, not all of the factors have the same possible scoring range, some carrying more weight than others. This is based on this researcher's own understanding of which factors should carrying more weight, and unfortunately, general intuition that a factor

such as source reliability is integral enough to warrant more ‘points’ on the table than another factor. Clearly the appropriateness of this weighting is up for debate and would need to stand up in the face of scientific testing and professional criticisms before any level of confidence could be placed in it. Second, some topics or situations may allow for certain levels of teamwork or use of numerous structured methods, thus at least partially negating those aspects of the table. Moreover, some analytic methods are more robust and have been found to be more beneficial to analyses than others, knowledge which analysts may not possess when scoring themselves on that factor. Finally, the table doesn’t leave any room for a ‘gut feeling,’ that many analysts rely on when they craft their forecasts. Even if all of the factors listed are present, or absent, an analyst may wish to modify his analytic confidence score based on some other characteristic not brought to light by the categories listed on the table. Though the easy fix for this problem would be to include a space for a narrative in which the analyst could express any such reservations or feelings, it does not address the absence of such a conveyance within the table itself.

Despite these and other possible initial criticisms, this researcher feels that the table described above and found in Appendix 10 is a step in the right direction, a step toward a more methodical and structured manner of rating analytic confidence based on appropriate and proven factors. This method would not only help reduce confusion amongst analysts and decision makers concerning analytic confidence and appropriate levels thereof, but would also begin the journey toward a more coherent way of teaching analysts what factors to consider when rating their own confidence. To use the same analogy this thesis began with, this method would not turn every confidence assessment

into a ‘slam dunk,’ but it will start getting each analyst to shoot the ball with the same, fundamentally appropriate, technique.

BIBLIOGRAPHY

Abbot, Sebastian. “CIRA and the Business of Revolutionizing Intelligence Analysis,” Good Harbor Report. http://www.goodharborreport.com/node/555.

Ahlfinger, Hailey L. “Confidence and Work Team Performance: A Study of the Groupthink Phenomenon,” 04/02/03. Clemson University. http://cujo.clemson.edu/manuscript.php?manuscript_ID=105.

Ahituv, Niv, Magid Igbaria, and Aviem Sella. “The Effects of Time Pressure and Completeness of Information on Decision Making.” *Journal of Management Information Systems* 15, no. 2 fall 1998: 153-172 [journal on-line]; available from http://jmis.bentley.edu/articles/v15_n2_p153/index.html; Internet; accessed 1 May 1507.

Amapedia. “Kahneman & Tversky,” 02/23/2006. Amazon.com. http://amapedia.amazon.com/view/Kahneman+&+Tversky/id=109612.

Branigin, William. “CIA Director Tenet Resigns,” 06/03/04. Washinton Post. http://www.washingtonpost.com/wp-dyn/articles/A12296-2004Jun3.html.

Carnegie Mellon University. “Baruch Fischhoff,” 2005. Carnegie Mellon University. http://sds.hss.cmu.edu/src/faculty/fischhoff.php.

Central Intelligence Agency. “Center for the Study of Intelligence,” 09/17/2007. Central Intelligence Agency. http://https://www.cia.gov/library/center-for-the-study-of-intelligence/index.html.

Central Intelligence Agency. “Transcript of C-SPAN Interview With CIA Director,” April 17, 2007. Central Intelligence Agency/CSPAN. http://https://www.cia.gov/news-information/press-releases-statements/press-release-archive-2007/transcript-of-c-span-interview-with-cia-director.html.

Chido, Diane E., and Richard M Seward Jr. *Structured Analysis Of Competing Hypotheses*. Mercyhurst College Erie, Pennsylvania: Mercyhurst College Institute of Intelligence Studies Press, 2006.

Cooper, Jeffrey R. “Curing Analytic Pathologies: Pathways to Improved Intelligence Analysis,” December, 2005. Central Intelligence Agency. http://https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/curing-analytic-pathologies-pathways-to-improved-intelligence-analysis-1/chapter_4.htm.

Davis, Jack. “Analytic Professionalism and the Policymaking Process: Q&A on a Challenging Relationship,” October, 2003. Central Intelligence Agency. http://https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/docs/v02n4p.htm.

Dougherty, Dr. Michael R.P. “What is Cognitive Psychology?,” University of Maryland. http://64.233.167.104/search?q=cache:IS9FuGJ8xDcJ:www.bsos.umd.edu/psyc/dougherty/classes/Psyc341/history%2520of%2520cognitive%2520psychology.ppt+http://www.bsos.umd.edu/psyc/dougherty/classes/Psyc341/history%2520of%2520cognitive%2520psychology.ppt&hl=en&ct=clnk&cd=1&gl=us.

Dunwoody, Philip T., Eric Haarbauer, Robert P. Mahan, Christopher Marino, and Chu-Chun Tang. “COGNITIVE ADAPTATION AND ITS CONSEQUENCES:A Test of Cognitive Continuum Theory.” *Journal of Behavioral Decision Making* 13, no. 1 Jan-Mar 2000: 35-54 [journal on-line]; available from http://faculty.juniata.edu/dunwoody/jbdm.pdf; Internet; accessed 15 May 2007.

Fischhoff, Baruch, and Don MacGregor. “Subjective Confidence in Forecasts.” *Journal of Forecasting* 1 1982: 155-172 [journal on-line]; available from http://http://stinet.dtic.mil/oai/oai?&verb=getRecord&metadataPrefix=html&identifier=ADA109730; Internet.

Heuer, Richards J. “Are Analysts Overconfident?*Unpublished Manuscript* (1980).

Heuer, Richards J. *Psychology of Intelligence Analysis*. Langley, Virginia: Central Intelligence Agency Center for the Study of Intelligence, 1999. Book on-line. Available from http://http://www.au.af.mil/au/awc/awcgate/psych-intel/index.html; Internet.

Janis, Irving. “What is Groupthink,” Psychologists for Social Responsibility. http://www.psysr.org/groupthink%20overview.htm.

Kent, Sherman. “Words of Estimative Probability,” Fall 1964. Central Intelligence Agency. http://https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/sherman-kent-and-the-board-of-national-estimates-collected-essays/6words.html#ft10.

Klayman, Joshua, Jack B. Soll, Claudia Gonzalez-Vallejo, and Sema Barlas. “Overconfidence: It Depends on How, What, and Whom You Ask.” *Organizational Behavior and Human Decision Processes* 79, no. 3 (1999): 216-247.

Koriat, Asher, Sarah Lichtenstein, and Baruch Fischoff, eds. “Reasons for Confidence.” *Journal of Experimental Psychology: Human Learning and Memory*, no. 6 (1980): 107-118.

Lahneman, William J. *The Future of Intelligence Analysis*. Volume 1 Final Report; University of Maryland, 2006. Book on-line. Available from

http://www.cissm.umd.edu/papers/files/future_intel_analysis_final_report1.pdf; Internet.

Lee, Michael D., and Matthew J. Dry. “Decision Making and Confidence Given Uncertain Advice.” *Cognitive Science* 30 03/08/06: 1081-1095 [journal on-line]; available from http://www.socsci.uci.edu/~mdlee/trust_lee_dry.pdf; Internet; accessed 15 May 2007.

Lyden, Michael. “The Efficacy of Accelerated Analysis in Strategic-Level Estimative Judgments.” (Master's Thesis, Mercyhurst College, May 2007): 63-70.

Mann, Leon, and Charlotte Tan. “The Hassled Decision Maker: The Effects of Perceived Time Pressure on Information Processing in Decision Making.” *Australian Journal of Management* 12/02/93: 197-210 [journal on-line]; available from http://www.agsm.unsw.edu.au/eajm/9312/pdf/mann.pdf; Internet; accessed 15 May 2007.

National Intelligence Council. “The Terrorist Threat to the US Homeland,” July, 2007. Office of the Director of National Intelligence. http://www.dni.gov/press_releases/20070717_release.pdf.

National Intelligence Council. “Iran: Nuclear Intentions and Capabilities,” 12/03/2007. Office of the Director of National Intelligence. http://www.dni.gov/press_releases/20071203_release.pdf.

Oskamp, Stuart. “Over Confidence in Case-Study Judgments.” *Journal of Consulting Psychology* 29, no. 3 (1965): 261-265.

“Overconfidence,” 2007. http://overconfidence.behaviouralfinance.net/.

Patterson, Emily S., Emilie M. Roth, and David D. Woods. “A Simulation Study of Computer-Supported Inferential Analysis Under Data Overload,” 12/01/99. The Ohio State University. http://csel.eng.ohio-state.edu/patterson//dissertation_tech_report.pdf.

Puncochar, Judith M., and Paul W. Fox. “Confidence in Individual and Group Decision Making: When "Two Heads" Are Worse Than One.” *Journal of Educational Psychology* 96, no. 3 2004: 582-591 [journal on-line]; available from http://www.psychologia.pl/~jasia/puncochar.pdf; Internet; accessed 18 May 2007.

Scholz, Roland W., Ralph Hansmann, Helmut W. Crott, and Harald A. Mieg. “Education in Environmental Planning: Effects of Group Discussion, Expert Information, and Case Study Participation on Judgment Accuracy,” November, 2001. Swiss Federal Institute of Technology, Zurich. http://e-collection.ethbib.ethz.ch/ecol-pool/incoll/incoll_453.pdf.

Schrage, Michael. “What Percent Is 'Slam Dunk'?,” 02/30/05. Washington Post. http://www.washingtonpost.com/wp-dyn/articles/A37115-2005Feb19.html.

Shahar, Yuval M.D., and Ph.D. “Confidence, Forecasting, Knowledge, and Calibration,” http://www.ise.bgu.ac.il/courses/dm/Lectures/lecture11.ppt.

Speier, Cheri. *Using Aggregated Data Under Time Pressure: A Mechanism for Coping with Information Overload*. University of Oklahoma, Michael F. Price College of Business, 1998. Book on-line. Available from http://csdl2.computer.org/comp/proceedings/hicss/1998/8236/02/82360004.pdf; Internet.

Speier, Cheri, Joseph S Valacich, and Iris Vessey. “The influence of task interruption on individual decision making: An information overload perspective.” *Decision Sciences* spring 1999 [journal on-line]; available from http://findarticles.com/p/articles/mi_qa3713/is_199904/ai_n8837238/pg_8; Internet; accessed 15 May 2007.

Surowiecki, James. *The Wisdom of Crowds*. New York: Doubleday, 2004.

Tetlock, Philip E. *Expert Political Judgment*. Princeton, New Jersey: Princeton University Press, 2005.

Tibrewal, Shradha, and John Poertner. “Confidence and Uncertainty in Casework Decisions: The Supervisor's Role,” School of Social Work, University of Illinois at Urbana-Champaign. http://cfrcwww.social.uiuc.edu/pubs/Pdf.files/confidence.pdf.

Topi, Heikki. “The effects of task complexity and time availability limitations on human performance in database query tasks. *International Journal of Human-Computer Studies* 62, no. 3 March 2005: 349-379 [journal on-line]; available from http://ezproxy.mercyhurst.edu/login?url=http://search.ebscohost.com.ezproxy.mercyhurst.edu/login.aspx?direct=true&db=aph&AN=17516901&site=ehost-live; Internet; accessed 15 May 2007.

Tsai, Claire, Joshua Klayman, and Reid Hastie. “Effects of Amount of Information on Judgment Accuracy and Confidence,” 2006. University of Chicago. http://home.uchicago.edu/~iwentsai/TsaiEtAl_Confidence.pdf.

United States Government. “Intelligence Reform and Terrorism Prevention Act of 2004,,” 12/17/04. http://www.nctc.gov/docs/pl108_458.pdf.

[Wapedia]. “Cognitive psychology,” 11/15/2007. Wapedia. http://wapedia.mobi/en/Cognitive_psychology.

Wesson, Caroline J. “The Communication and Influence of Confidence and Uncertainty,” November, 2005. University of Leicester. http://www.le.ac.uk/pc/bdp5/Cari's%20Thesis.pdf.

Wheaton, Kristan. “Part 5 -- Enough Exposition! Let’s Get Down To It…,” January 8, 2008. Kristan Wheaton. http://sourcesandmethods.blogspot.com/.

APPENDICES

Appendix 1: Email Announcing Experiment & Soliciting Participants

Attention: Sophs, Juniors, Seniors, G1s & G2s

I will be conducting an experiment for my graduate thesis on the evening of Monday December 10th in Room 117 of the Wayne Street building.

The following professors are offering extra credit (for Winter Term Intel Courses) to any of their students who participate in the experiment: Wheaton, Mills, Wozneak, Grabelski, Mulligan, Welch & Heibel

If you are in a class with an Intel professor other than those listed above, you are still more than welcome to participate in the experiment.

There are 3 time slots available for you to sign up for: 1800-1900, 1915-2015, 2030-2130. Each session will last approximately 1 hour. The sign-up sheet is posted on the bulletin board next to Room 117 (across from Professor Heibel's office). Please sign up for whichever session has an empty slot still available. If one time slot has filled, please choose one that still has empty places.

My thesis experiment is focused on analytic confidence in intelligence analysis. The experiment will not be difficult, just a simple analysis of an intelligence-oriented scenario. No other preparation is needed to participate, just sign-up and be present at whichever time you picked.

Thank you.

Please contact me via E-mail with any questions or concerns:
Josh Peterson (G2)
jpeter30@mercyhurst.edu

Appendix 2: Flyer Announcing Experiment & Soliciting Participants

What: Graduate Thesis Experiment

Why: Extra Credit For Intel Courses

- (Professors: Wozneak, Grabelski, Mills, Mulligan, Wheaton, Hiebel, and Welch are giving extra credit for participation)
- If you are in an Intel class with a professor who is not listed, you are still welcome to participate.

Where: Room 117 in the Intel building

Who: Open to Soph, Jr, Sr, G1, G2

When: Monday December 10th

- Sign up for 1 of 3 time slots
  - 1800-1900
  - 1915-2015
  - 2030-2130
- Sign-up sheet is on the bulletin board next to Room #117

Each session should last approximately 1 hour total. My thesis experiment is focused on analytic confidence in intelligence analysis. The experiment will not be difficult, just a simple analysis of an intelligence-oriented scenario. No other preparation is needed to participate, just sign-up and be present at whichever time you picked. If one time slot has filled, please choose one that still has empty places.

Thank you. Please contact me with any questions or concerns.
-Josh Peterson (G2)
jpeter30@mercyhurst.edu

Appendix 3: Experiment Sign-Up Sheets

Graduate Thesis Experiment Sign-Up

Please fill in your name in an empty slot.

1800-1900:

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____
19. _____
20. _____
21. _____

1915-2015:

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____
19. _____
20. _____
21. _____

2030-2130

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____
19. _____
20. _____
21. _____

Please arrive a few minutes before your session is scheduled to start. If there are any questions or concerns, please send me an email at jpeter30@mercyhurst.edu .

Thank you.
-Josh Peterson

Appendix 4: Institutional Review Board Proposal Form

Date Submitted: *9/28/2007*	Advisor's Name (if applicable): *Kristan Wheaton*
Investigator(s): *Josh Peterson*	Advisor's E-mail: *kwheaton@mercyhurst.edu*
Investigator Address: *805 E. Grandview #104* *Erie, PA 16504*	Advisor's Signature of Approval: [*X*] Place X here if advisor has approved research
Investigator(s) E-mail: *Jpeter30@mercyhurst.edu*	Title of Research Project: *Appropriate Measures of Analytic Confidence in Intelligence Analysis*
Investigator Telephone Number: *612-868-3066*	Date of Initial Data Collection: *TBD, anticipate October-December 2007*

Please describe the proposed research and its purpose, in narrative form:

Analytic confidence is an expression of uncertainty in intelligence analysis meant to convey an analyst's confidence in his/her forecast to the decisionmaker reading it. Unfortunately, very little information is available on what that expression of uncertainty should be based on, despite a Congressional mandate to clearly convey analytic confidence in finished intelligence products of the United States Intelligence Community. Currently, analysts intuitively construct an expression of analytic confidence, which is often poorly calibrated. Even in our nation's National Intelligence Estimates, analytic confidence gets scantly more than a passing mention and is confusing in its application.

The purpose of this study is to assess what are appropriate measures to consider when constructing an analytic confidence rating, and in doing so go from a flawed intuitive-based guess to a method of developing an appropriate confidence expression. I have developed a list of factors I believe to be appropriate measures of analytic confidence, and taken together, I believe they will demonstrate a more reliable method of constructing analytic confidence ratings in intelligence analysis. I plan to test my method using intelligence scenarios and intelligence analysts (both graduate and undergraduate students in the intelligence program at Mercyhurst) to determine if it truly is a viable system analysts can use to express analytic confidence.

Indicate the materials, techniques, and procedures to be used **(submit copies of materials):**
Materials:

Exercise Scenarios
Writing Utensils
Post-Test Questionnaire

Procedure:

One week prior to the study I will send out reminders to those who have volunteered to participate. I will send out another reminder the day before the study. During the actual study I will begin by going over the directions of the exercise and explaining what is expected of the participants and how they will get credit for their participation. After the introduction, I will pass out the study's materials and instruct the students to begin their analyses. Following the completion of the exercise, I will ask the participants to fill-out a questionnaire (attached at end) and provide feedback regarding both the topic and the experiment.

I plan to conduct my experiment three times, on three different nights. They will vary only in which handout I give them. There will be 2 experimental groups and a control group. All groups will be given the exact same scenario; I just will vary some information which will after my operational variable: analytic confidence level. (Please see the forms below)

1. Do you have **external funding** for this research (money coming from outside the College)? Yes[ ] No[*X*]

Funding Source (if applicable):

2. Will the participants in your study come from a **population requiring special protection;** in other words, are your subjects someone other than Mercyhurst College students (i.e., children 17-years-old or younger, elderly, criminals, welfare recipients, persons with disabilities, NCAA athletes)? Yes[ ] No[*X*]

If your participants include a population requiring special protection, describe how you will obtain consent from their legal guardians and/or from them directly to insure their full and free consent to participate.
N/A

Indicate the approximate number of participants, the source of the participant pool, and recruitment procedures for your research:
I plan to have approximately 100 participants. I plan to recruit undergraduate and graduate students in the intelligence studies department through a department-wide email. I will select the students on a first come, first serve basis.

Will participants receive any payment or compensation for their participation in your research *(this includes money, gifts, extra credit, etc.)*? Yes[*X*] No[ ]

If yes, please explain: *Hopefully extra credit for participation. I am not sure if all of the intelligence professors will give extra credit, but in the past they have all been willing to grant it for participating in an experiment.*

3. Will the participants in your study be at any physical or psychological **risk** (risk is defined as any procedure that is invasive to the body, such as injections or drawing blood; any procedure that may cause undue fatigue; any procedure that may be of a sensitive nature, such as asking questions about sexual behaviors or practices) such that participants could be emotionally or mentally upset? Yes[ ] No[*X*]

Describe any harmful effects and/or risks to the participants' health, safety, and emotional or social well being, incurred as a result of participating in this research, and how you will insure that these risks will be mitigated:
None.

4. Will the participants in your study be **deceived** in any way while participating in this research? Yes[ ] No[*X*]

If your research makes use of any deception of the respondents, state what other alternative (e.g., non-deceptive) procedures were considered and why they weren't chosen:
N/A

5. Will you have a written **informed consent** form for participants to sign, and will you have appropriate **debriefing** arrangements in place? Yes[*X*] No[ ]

Describe how participants will be clearly and completely informed of the true nature and purpose of the research, whether deception is involved or not **(submit informed consent form and debriefing statement):**
Prior to the training sessions, participants will be provided with a general overview of what will occur during the session as well as the consent form, which will also describe what is expected of them. Following the administrative questionnaire, participants will be provided with a debriefing statement that will explain how the results from the session will be used (please see forms at the end of this proposal).

Please include the following statement at the bottom of your informed consent form: "Research at Mercyhurst College which involves human participants is overseen by the Institutional Review Board. Questions or problems regarding your rights as a participant should be addressed to Mr. Tim Harvey Institutional Review Board Chair; Mercyhurst College; 501 East 38th Street; Erie, Pennsylvania 16546-0001; Telephone (814) 824-3372."

6. Describe the nature of the data you will collect and your procedures for insuring that **confidentiality** is maintained, both in the record keeping and presentation of this data:
Names are not required for my research and thus no names will be used in the recording of the results or the presentation of my data. Names will only be used to notify professors of participation in order for them to correctly assign extra credit.

7. Identify the potential **benefits** of this research on research participants and humankind in general.

Potential benefits include:
For participants:
An opportunity to practice the intelligence analysis skills they have learned in the classroom in an experiment aimed at clarifying what analytic confidence ratings should be based on. Students are often asked to assess their analytic confidence in assignments in Intelligence Studies courses, and this experiment hopes to better understand how those ratings should be calculated.

For the Intelligence Community:
Currently the United States Intelligence Community bases their analytic confidence nearly exclusively on source reliability. While this may be acceptable, it is not optimal. This experiment hopes to prove that more factors should be considered, and that in doing so the Community's analytic confidence ratings will be better calibrated, and thus more informative for our nation's decisionmakers.

Please submit this file and accompanying materials to the IRB Chair, Tim Harvey, via electronic mail (tharvey@mercyhurst.edu) for review.

Appendix 5: Consent Form

Analytic Confidence
Participation Consent Form

The purpose of this research is to test a method to determine appropriate factors to consider when formulating an analytic confidence rating in intelligence analysis.

Your participation involves a short instruction period, completing two short intelligence analyses, and filling out a questionnaire. This process should take no longer than 1 hour. Your name WILL NOT appear in any information disseminated by the researcher. Your name will only be used to notify professors of your participation in order for them to assign extra credit.

There are no foreseeable risks or discomforts associated with your participation in this study. Participation is voluntary and you have the right to opt out of the study at any time for any reason without penalty.

I, _____, acknowledge that my involvement in this research is voluntary and agree to submit my data for the purpose of this research.

_____ _____

Signature *Date*

_____ _____

Printed Name *Class*

Name(s) of professors offering extra credit: _____

Researcher's Signature: _____

If you have any further question about Analytic Confidence or this research you can contact me at jpeter30@mercyhurst.edu

Research at Mercyhurst College which involves human participants is overseen by the Institutional Review Board. Questions or problems regarding your rights as a participant should be addressed to Tim Harvey; Institutional Review Board Chair;

Mercyhurst College; 501 East 38th Street; Erie, Pennsylvania 16546-0001; Telephone (814) 824-3372. tharvey@mercyhurst.edu

Josh Peterson, *Applied Intelligence Master's Student, Mercyhurst College* 612-868-3066
Kristan Wheaton, *Research Advisor, Mercyhurst College* 814-824-3023

Appendix 6: Experiment Scenario Handouts

EXPERIMENT SECTION #1 (*HIGH CONFIDENCE)*

You are a national security advisor to the leader of your country, Country X. Your leader has tasked you to give him an estimate on the intentions of Country Y (your neighboring country) in light of recent events. You leader wants to know if an attack by country Y is likely and what your level of analytic confidence is in that decision. Your country's armed forces are small but capable of defending the country; mobilization will begin if your analysis indicates likely hostile intentions on the part of Country Y.

The information you currently have available is found below. (Source Reliability)

- The two countries have fought 4 wars in the past 20 years, the last war ended in a treaty settling disputed territory, the terms of which slightly favored Country X and embittered Country Y. (10)
- Media sources report that Country Y has called its military officers back to bases, and has cancelled leave to the soldiers. (8.5)
- Imagery suggests an unusually high concentration of military forces near your shared border. Country Y's official explanation for this build-up is military exercises; this build-up appears identical to operations before previous wars erupted. (9.5)
- Radio transmissions from Country Y's aircraft carrier battle group have ceased. Satellite imagery is unavailable, and airborne attempts to locate it have been unsuccessful. (10)
- Country Y's air defense systems are rumored to be grossly inadequate to defend the country in the event of a conflict; the last war was won due to your country's air superiority. (5)
- A spy reported that Country Y's soldiers have been issued new desert uniforms, which blend better with the border region's environment; Country Y is only 2% desert. (8.5)
- 2 of your other spies within Country Y's armed forces have been mysteriously out of contact over the past 6 months. (8.5)
- Your MASINT intelligence agency has suddenly lost track of County Y's submarine fleet operating in international waters off the coast, near the harbor where your primary naval fleet is docked. (9)
- Hurricane season will begin in approximately 4 months. Weather forecasters predict that if a hurricane hits, transportation of any kind between Country Y's capitol and the shared border region could become difficult. (3)
- One of your diplomats overheard what he thought to be a conversation in which Country Y's ambassador to Country X discussed plans for evacuating back to his native country with his staff. (9.5)
- Reports of a widespread outbreak of West Nile virus within the elite armored divisions of Country Y's army have surfaced. (2)

- Part of your army's General Staff may be abroad for the next two weeks to attend an international arms conference, while reports indicate Country Y's General Staff may be cancelling their visit to the same conference. (9)
- A reconnaissance drone was shot down, while flying over one of your country's missile defense batteries deep in your territory, the model of the drone may be the same as those used by Country Y's air force. (9.5)
- Country Y's air defense budget is rumored to have grown over 35% in the past two years. (9)
- Your country is celebrating a national holiday next weekend and much of your armed forces have been granted leave. (10)

Please use the computer in front of you to fill out an Analysis of Competing Hypotheses given the information presented above. Once you have completed your ACH please form into your teams and discuss your AHC and your analysis amongst the team. Regardless of whether or not a consensus is reached, please write your own analysis, citing the evidence you personally have based it on, and then rate your analytic confidence in your analysis.

Important Information:

Source Reliability:

Source Reliability reflects the accuracy and reliability of a particular source over time. Sources with high reliability have been proven to have produced accurate, consistently reliable, intelligence in the past. Sources with low reliability lack the accuracy and proven track record commensurate with more reliable sources.

- 1-10 point scale conveying the reliability of the sources used for that piece of intelligence/report, 1 being the lowest reliability and 10 the highest.

Analytic Confidence:

Analytic Confidence reflects the level of confidence an analyst has in his or her estimates, analyses, and the methods in which they produced them. It is not the same as using words of estimative probability, which indicate likelihood. It is possible for an analyst to suggest an event is virtually certain based on the available evidence, yet have a low amount of confidence in that forecast due to a variety of factors or vice versa.

- Mark your analytic confidence on the scale at the bottom of the next page. Please see the board at the front of the room for an example.

Important Note:

There is a distinction between psychological confidence and analytic confidence. *Psychological* confidence in this case is how confident you 'feel' about something, which is based not based on any system or scientific process. An example would be a student feeling

confident their collegiate hockey team will win the national championship after going to only 1 game and knowing nothing about hockey. *Analytic* confidence is 'legitimate' confidence derived from on the actual analysis and the analytic process. Concepts you might consider when assessing your analytic confidence may include: your subject matter expertise, source reliability, time spent on the analysis, collaboration with others, source corroboration, use of structured methods, and the complexity of the analysis.

Please write your analysis in the space provided below.

It is likely that _____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Please mark your analytic confidence (see example on the board).

Low |----------| High

EXPERIMENT SECTION #2 (LOW CONFIDENCE)

You are a national security advisor to the leader of your country, Country X. Your leader has tasked you to give him an estimate on the intentions of Country Y (your neighboring country) in light of recent events. You leader wants to know if an attack by country Y is likely and what your level of analytic confidence is in that decision. Your country's armed forces are small but capable of defending the country; mobilization will begin if your analysis indicates likely hostile intentions on the part of Country Y. You have 15 minutes to complete your analysis, write your analysis in the space provided.

The information you currently have available is found below. (Source Reliability)

- The two countries have fought 4 wars in the past 20 years, the last war ended in a treaty settling disputed territory, the terms of which slightly favored Country X and embittered Country Y. (5)
- Media sources report that Country Y has called its military officers back to bases, and has cancelled leave to the soldiers. (3.5)
- Imagery suggests an unusually high concentration of military forces near your shared border. Country Y's official explanation for this build-up is military exercises; this build-up appears identical to operations before previous wars erupted. (4.5)
- Radio transmissions from Country Y's aircraft carrier battle group have ceased. Satellite imagery is unavailable, and airborne attempts to locate it have been unsuccessful. (4)
- Country Y's air defense systems are rumored to be grossly inadequate to defend the country in the event of a conflict; the last war was won due to your country's air superiority. (5)
- A spy reported that Country Y's soldiers have been issued new desert uniforms, which blend better with the border region's environment; Country Y is only 2% desert. (3.5)
- 2 of your other spies within Country Y's armed forces have been mysteriously out of contact over the past 6 months. (3.5)
- Your MASINT intelligence agency has suddenly lost track of County Y's submarine fleet operating in international waters off the coast, near the harbor where your primary naval fleet is docked. (4)
- Hurricane season will begin in approximately 4 months. Weather forecasters predict that if a hurricane hits, transportation of any kind between Country Y's capitol and the shared border region could become difficult. (3)
- One of your diplomats overheard what he thought to be a conversation in which Country Y's ambassador to Country X discussed plans for evacuating back to his native country with his staff. (4.5)
- Reports of a widespread outbreak of West Nile virus within the elite armored divisions of Country Y's army have surfaced. (2)

- Part of your army's General Staff may be abroad for the next two weeks to attend an international arms conference, while reports indicate Country Y's General Staff may be cancelling their visit to the same conference. (4)
- A reconnaissance drone was shot down, while flying over one of your country's missile defense batteries deep in your territory, the model of the drone may be the same as those used by Country Y's air force. (4.5)
- Country Y's air defense budget is rumored to have grown over 35% in the past two years. (4)
- Your country is celebrating a national holiday next weekend and much of your armed forces have been granted leave. (5)

Please use the attached paper for your analysis of the information presented above. Additionally, please cite the evidence you have based your analysis on, and then rate your analytic confidence in your analysis.

Important Information:

Source Reliability:

Source Reliability reflects the accuracy and reliability of a particular source over time. Sources with high reliability have been proven to have produced accurate, consistently reliable, intelligence in the past. Sources with low reliability lack the accuracy and proven track record commensurate with more reliable sources.

- 1-10 point scale conveying the reliability of the sources used for that piece of intelligence/report, 1 being the lowest reliability and 10 the highest.

Analytic Confidence:

Analytic Confidence reflects the level of confidence an analyst has in his or her estimates and analyses. It is not the same as using words of estimative probability, which indicate likelihood. It is possible for an analyst to suggest an event is virtually certain based on the available evidence, yet have a low amount of confidence in that forecast due to a variety of factors or vice versa.

- Mark your analytic confidence on the scale at the bottom of the next page. Please see the board at the front of the room for an example.

Important Note:

There is a distinction between psychological confidence and analytic confidence. *Psychological* confidence in this case is how confident you 'feel' about something, which is based not based on any system or scientific process. An example would be a student feeling confident their collegiate hockey team will win the national championship after going to only 1 game and knowing nothing about hockey. *Analytic* confidence is 'legitimate' confidence

derived from on the actual analysis and the analytic process. Concepts you might consider when assessing your analytic confidence may include: your subject matter expertise, source reliability, time spent on the analysis, collaboration with others, source corroboration, use of structured methods, and the complexity of the analysis.

Please write your analysis in the space provided below.

It is likely that _____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Please mark your analytic confidence (see example on the board).

Low |----------| High

EXPERIMENT SECTION #3 (CONTROL GROUP)

You are a national security advisor to the leader of your country, Country X. Your leader has tasked you to give him an estimate on the intentions of Country Y (your neighboring country) in light of recent events. You leader wants to know if an attack by country Y is likely and what your level of analytic confidence is in that decision. Your country's armed forces are small but capable of defending the country; mobilization will begin if your analysis indicates likely hostile intentions on the part of Country Y.

The information you currently have available is found below.

- The two countries have fought 4 wars in the past 20 years, the last war ended in a treaty settling disputed territory, the terms of which slightly favored Country X and embittered Country Y.
- Media sources report that Country Y has called its military officers back to bases, and has cancelled leave to the soldiers.
- Imagery suggests an unusually high concentration of military forces near your shared border. Country Y's official explanation for this build-up is military exercises; this build-up appears identical to operations before previous wars erupted.
- Radio transmissions from Country Y's aircraft carrier battle group have ceased. Satellite imagery is unavailable, and airborne attempts to locate it have been unsuccessful.
- Country Y's air defense systems are rumored to be grossly inadequate to defend the country in the event of a conflict; the last war was won due to your country's air superiority.
- A spy reported that Country Y's soldiers have been issued new desert uniforms, which blend better with the border region's environment; Country Y is only 2% desert.
- 2 of your other spies within Country Y's armed forces have been mysteriously out of contact over the past 6 months.
- Your MASINT intelligence agency has suddenly lost track of County Y's submarine fleet operating in international waters off the coast, near the harbor where your primary naval fleet is docked.
- Hurricane season will begin in approximately 4 months. Weather forecasters predict that if a hurricane hits, transportation of any kind between Country Y's capitol and the shared border region could become difficult.
- One of your diplomats overheard what he thought to be a conversation in which Country Y's ambassador to Country X discussed plans for evacuating back to his native country with his staff.
- Reports of a widespread outbreak of West Nile virus within the elite armored divisions of Country Y's army have surfaced.

- Part of your army's General Staff may be abroad for the next two weeks to attend an international arms conference, while reports indicate Country Y's General Staff may be cancelling their visit to the same conference.
- A reconnaissance drone was shot down, while flying over one of your country's missile defense batteries deep in your territory, the model of the drone may be the same as those used by Country Y's air force.
- Country Y's air defense budget is rumored to have grown over 35% in the past two years.
- Your country is celebrating a national holiday next weekend and much of your armed forces have been granted leave.

Please use the attached paper for your analysis of the information presented above. Additionally, please cite the evidence you have based your analysis on, your overall rating of the sources' reliability, and then rate your analytic confidence in your analysis.

Important Information:

Analytic Confidence:

Analytic Confidence reflects the level of confidence an analyst has in his or her estimates and analyses. It is not the same as using words of estimative probability, which indicate likelihood. It is possible for an analyst to suggest an event is virtually certain based on the available evidence, yet have a low amount of confidence in that forecast due to a variety of factors or vice versa.

o Mark your analytic confidence on the scale at the bottom of the next page. Please see the board at the front of the room for an example.

Important Note:

There is a distinction between psychological confidence and analytic confidence. *Psychological* confidence in this case is how confident you 'feel' about something, which is based not based on any system or scientific process. An example would be a student feeling confident their collegiate hockey team will win the national championship after going to only 1 game and knowing nothing about hockey. *Analytic* confidence is 'legitimate' confidence derived from on the actual analysis and the analytic process. Concepts you might consider when assessing your analytic confidence may include: your subject matter expertise, source reliability, time spent on the analysis, collaboration with others, source corroboration, use of structured methods, and the complexity of the analysis.

Please write your analysis in the space provided below.

It is likely that _____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Please mark your analytic confidence (see example on the board).

Low |----------| High

Appendix 7: Debriefing Sheet

Analytic Confidence
Participation Debriefing

Thank you for participating in this research process. I appreciate your contribution and willingness to support the student research process.

The purpose of this study was to determine appropriate factors to consider when formulating a statement of analytic confidence in intelligence analysis. Currently there has been little research done on this topic, and this study hopes to take the first of many steps in moving beyond intuitive reasoning or reliance only one factor and toward a method of better calibrating analytic confidence. My experiments today were designed to focus on factors I have found to be correlated with well calibrated confidence, and thus demonstrate their significance in assessing analytic confidence.

Improved analytic confidence and clarity in conveying it have been requested by those at the highest level of our government. Congress has even tasked the Intelligence Community to incorporate this vital information in its analyses. Recent National Intelligence Estimates have begun to address the dearth of coverage on analytic confidence. I plan to use the results from this study to support my assertion that current practices in formulating analytic confidence in intelligence analysis are not optimal, and that more factors should be considered in order to have better calibrated analytic confidence ratings.

If you have any further question about analytic confidence or this research you can contact me at jpeter30@mercyhurst.edu.

Appendix 8: Detailed Discussion of Structure Analysis of Competing Hypotheses

The follow excerpts were taken from Richards Heuer's book The Psychology of Intelligence Analysis, chapter 8 'Analysis of Competing Hypotheses.'

> Analysis of competing hypotheses, sometimes abbreviated ACH, is a tool to aid judgment on important issues requiring careful weighing of alternative explanations or conclusions. It helps an analyst overcome, or at least minimize, some of the cognitive limitations that make prescient intelligence analysis so difficult to achieve.
>
> ACH is an eight-step procedure grounded in basic insights from cognitive psychology, decision analysis, and the scientific method. It is a surprisingly effective, proven process that helps analysts avoid common analytic pitfalls. Because of its thoroughness, it is particularly appropriate for controversial issues when analysts want to leave an audit trail to show what they considered and how they arrived at their judgment.
>
> Analysis of competing hypotheses (ACH) requires an analyst to explicitly identify all the reasonable alternatives and have them compete against each other for the analyst's favor, rather than evaluating their plausibility one at a time.
>
> The way most analysts go about their business is to pick out what they suspect intuitively is the most likely answer, then look at the available information from the point of view of whether or not it supports this answer. If the evidence seems to support the favorite hypothesis, analysts pat themselves on the back ("See, I knew it all along!") and look no further. If it does not, they either reject the evidence as misleading or develop another hypothesis and go through the same procedure again. Decision analysts call this a satisficing strategy. (See Chapter 4, Strategies for Analytical Judgment.) Satisficing means picking the first solution that seems satisfactory, rather than going through all the possibilities to identify the very best solution. There may be several seemingly satisfactory solutions, but there is only one best solution.
>
> Chapter 4 discussed the weaknesses in this approach. The principal concern is that if analysts focus mainly on trying to confirm one hypothesis they think is probably true, they can easily be led astray by the fact that there is so much evidence to support their point of view. They fail to recognize that most of this evidence is also consistent with other explanations or conclusions, and that these other alternatives have not been refuted.

Simultaneous evaluation of multiple, competing hypotheses is very difficult to do. To retain three to five or even seven hypotheses in working memory and note how each item of information fits into each hypothesis is beyond the mental capabilities of most people. It takes far greater mental agility than listing evidence supporting a single hypothesis that was pre-judged as the most likely answer. It can be accomplished, though, with the help of the simple procedures discussed here. The box below contains a step-by-step outline of the ACH process.

Step-by-Step Outline of Analysis of Competing Hypotheses

1. Identify the possible hypotheses to be considered. Use a group of analysts with different perspectives to brainstorm the possibilities.

2. Make a list of significant evidence and arguments for and against each hypothesis.

3. Prepare a matrix with hypotheses across the top and evidence down the side. Analyze the "diagnosticity" of the evidence and arguments--that is, identify which items are most helpful in judging the relative likelihood of the hypotheses.

4. Refine the matrix. Reconsider the hypotheses and delete evidence and arguments that have no diagnostic value.

5. Draw tentative conclusions about the relative likelihood of each hypothesis. Proceed by trying to disprove the hypotheses rather than prove them.

6. Analyze how sensitive your conclusion is to a few critical items of evidence. Consider the consequences for your analysis if that evidence were wrong, misleading, or subject to a different interpretation.

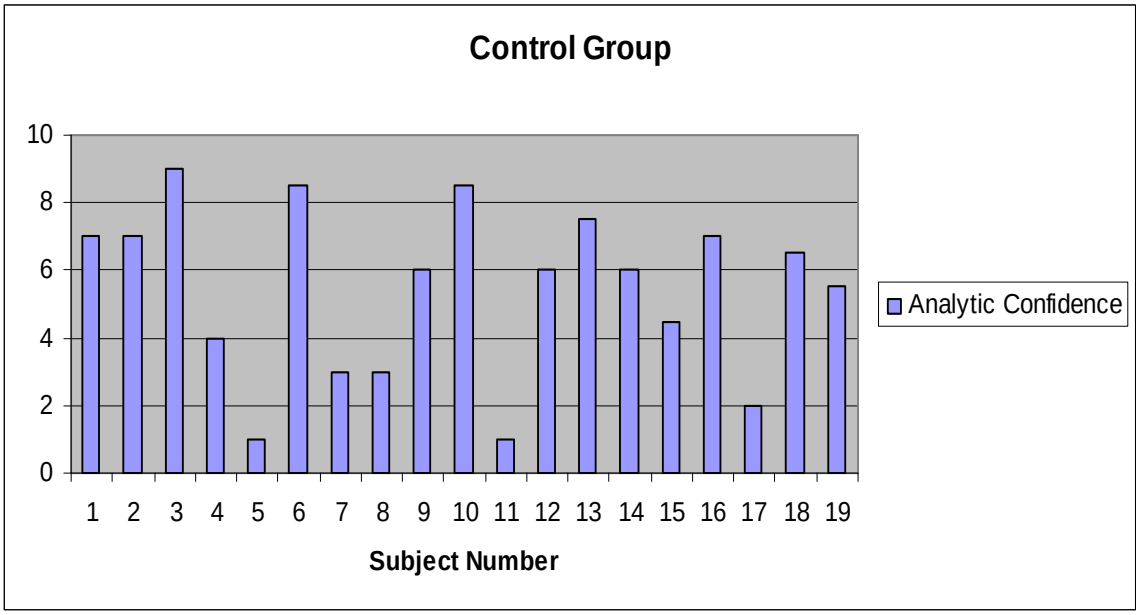
7. Report conclusions. Discuss the relative likelihood of all the hypotheses, not just the most likely one.

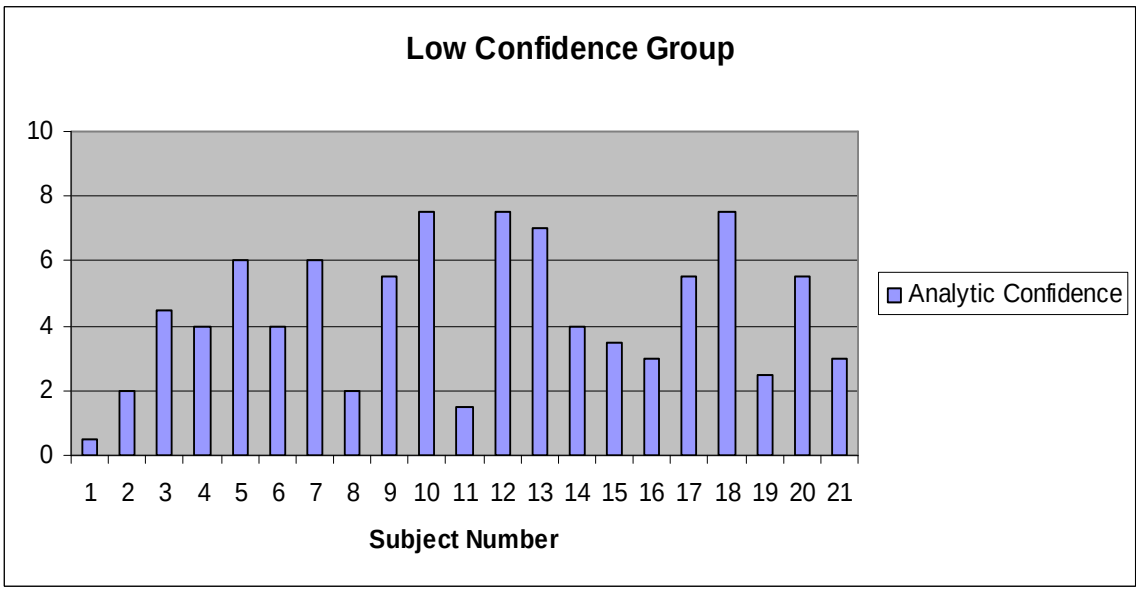
8. Identify milestones for future observation that may indicate events are taking a different course than expected.

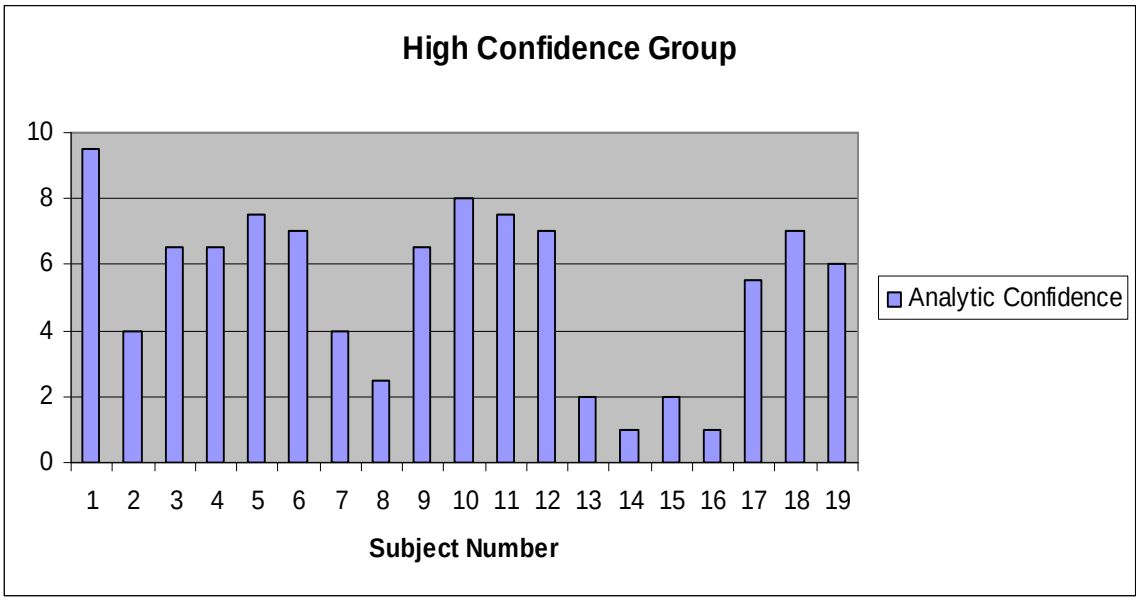
For more an even more detailed discussion/explanation of analysis of competing hypotheses, please refer to Heuer’s book online at:

http://www.au.af.mil/au/awc/awcgate/psych-intel/art11.html. Software to perform this type of analysis is available for free download at: http://www2.parc.com/istl/projects/ach/ach.html.

Appendix 9 Misc. Charts From Experiment:







Statistical Breakdown of Experimental Group's Results From SPSS

Descriptives

Analytic Confidence

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
High	15	6.3333	1.75933	.45426	5.3590	7.3076	2.50	9.50
Low	21	4.4048	2.10130	.45854	3.4483	5.3613	.50	7.50
Control	19	5.4211	2.48475	.57004	4.2234	6.6187	1.00	9.00
Total	55	5.2818	2.26029	.30478	4.6708	5.8929	.50	9.50

Appendix 10 – Analytic Confidence Assessment Tool

PETERSON TABLE OF ANALYTIC CONFIDENCE ASSESSMENT	Points Possible	Example Points
Use of Structured Method(s) In Analysis	**(1-10)**	***7***
For example: ACH, IPB, Social Networking, Bayes, Simulation, etc…		
10 indicating highest possible score when considering factors below		
Consider:		
Number of methods used		
Applicability of methods to the analysis		
Level of robustness of method		
Degree to which methods' results coincide		
Overall Source Reliability	**(1-10)**	***7***
A rating of 10 indicates the highest reliability		
Source Corroboration/Agreement: *Level of conflict amongst sources*	**(1-5)**	***4***
5: No confliction amongst sources		
4: Very little conflict amongst sources		
3: Moderate conflict amongst sources		
2: Significant conflict amongst sources		
1: Sources conflict on nearly all points		
Level Of Expertise On Subject/Topic & Experience	**(1-5)**	***2***
5: Deep, intimate knowledge and understanding & 3+ years experience with topic		
4: Wide knowledge & 1-3 years experience with topic		
3: Moderate knowledge & 6-12 months experience with topic		
2: Minimal knowledge & 0-5 months experience with the topic		
1: No knowledge & no experience with the topic		
Amount of Collaboration:	**(1-5)**	***2***
5: Part of aggregated individual analyses		
4: Worked on a team		
3: Worked with a partner		
2: Casual discussion		
1: Completely individual work		
Task Complexity	**(1-5)**	***3***
5: Minimally complex & challenging		
4: Somewhat complex & challenging		
3: Moderately complex & challenging		
2: Quite complex & challenging		
1: Very complex & highly challenging		
Time Pressure: *Time given to make analysis*	**(1-5)**	***4***
5: No deadline		
4: Easy to meet deadline		
3: Moderate deadline		
2: Demanding deadline		
1: Grossly inadequate deadline		
	Score:	**29**
	Total Possible:	**45**
	Score:	**0.644444444**
		x 10
	Analytic Confidence Adjusted Score:	**6.4**