Bayesian Reasoning Method For Intelligence Using Natural Frequencies

JENNIFER LYNN LEE

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
MAY 2007

DEPARTMENT OF INTELLIGENCE STUDIES
MERCYHURST COLLEGE
ERIE, PENNSYLVANIA

Bayesian Reasoning Method For Intelligence Using Natural Frequencies

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:

JENNIFER LYNN LEE

Certificate of Approval:

____
Kristan J. Wheaton
Assistant Professor
Department of Intelligence Studies

____
James G. Breckenridge
Chairman/Assistant Professor
Department of Intelligence Studies

____
Barbara A. Behan
Vice President
Office of Academic Affairs

MAY 2007

DEDICATION

This work is dedicated to my friends and family who have supported me throughout this tireless process.

ACKNOWLEDGEMENTS

I would like to thank Professor Kristan Wheaton for his guidance, advice, and support throughout the thesis process. I would also like to thank Professor Hemangini Deshmukh for her help analyzing my findings and performing the statistical analysis, which I would have not been able to do on my own. Finally, I would like to thank my fellow classmates and friends for their help with the research and the editing of this thesis.

ABSTRACT OF THE THESIS

Bayesian Reasoning Method For Intelligence Using Natural Frequencies

By

Jennifer Lynn Lee

Master of Science in Applied Intelligence

Mercyhurst College, 2007

Professor Kristan J. Wheaton, Chair

This thesis introduces a paper and pencil natural frequency tree method for intelligence analysts to use to perform Bayesian reasoning. This method is based on Peter Sedlmeier and Gerd Gigerenzer's computerized natural frequency tree training program, which shows that the use of natural frequencies in structured methods improves people's abilities to perform Bayesian reasoning. [1] Through the use of three different experiments, the author tests the method on its ability to improve Bayesian reasoning performance when compared to wording Bayesian problems using natural frequencies. The findings indicate the paper and pencil method introduced in this thesis leads to a high level of performance when taught to future intelligence analysts who are unfamiliar with Bayesian reasoning. Additionally, this study shows that it is possible to teach intelligence analysts a new, structured method in a small amount of time. This thesis can be considered a starting point for the vast amount of research that can be done on the benefits of using Bayesian reasoning in the intelligence field.

[1] Sedlmeier, Peter, and Gerd Gigerenzer. "Teaching Bayesian Reasoning in Less Than Two Hours," Journal *of Experimental Psychology: General* 130 (2001): 380-400.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE:

INTRODUCTION

On December 17, 2004, President George W. Bush signed into law the Intelligence Reform and Terrorism Prevention Act, the most dramatic national intelligence capabilities reform since the signing of the National Security Act of 1947 by then President Harry S. Truman. Based on suggestions from the 9/11 Commission Report and the WMD Commission Report (Report of the Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction), this new law calls for better cooperation throughout the intelligence community and better analytical integrity. Specifically, the Act calls for the use of analytic methodologies, tradecraft, and practices that provide timely and objective intelligence.[2]

One method that has been discussed in the intelligence community as a possible analytic technique meeting these requirements is Bayesian reasoning. Bayesian reasoning is a statistical technique based on the scientific method that combines base rates with evidence to help decision makers make accurate decisions. It does this by eliminating the effects of biases and allowing decision makers to see the quantitative impact evidence has on the chance of an event happening or not happening. Since Bayesian reasoning encompasses both possibilities, it appears to be an excellent methodology for the intelligence community.

Additionally, it is argued that Bayesian reasoning can help reduce the effects of cognitive biases. Cognitive biases result from simplified information processing

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

strategies and can lead to errors in judgments. Cognitive biases can easily impact intelligence analysis by "affect[ing] the evaluation of evidence, perception of cause and effect, estimation of probabilities, and retrospective evaluation of intelligence reports."[3]

However, for most people, Bayesian reasoning is a difficult concept to understand, and this is not surprising since most people cringe at the thought of performing math, especially math that involves the use of statistics. In fact, there are multiple studies that claim people are incapable of using Bayesian inference and that even formal training on Bayes' Rule is not able to solve this problem. According to them, people's "minds are not built (for whatever reason) to work by the rules of probability."[4]

Nevertheless, German psychologists Ulrich Hoffrage and Gerd Gigerenzer argue that there is an easier way to learn Bayesian reasoning through what they call "natural frequencies." Simply put, natural frequencies put statistical information into a form that people without strong mathematical backgrounds can easily understand. The two psychologists argue that the human mind is not attuned to statistical information, but if the information is put into a form that humans understand, then Bayesian reasoning becomes simpler.[5]

Building off of this idea, Peter Sedlmeier and Gigerenzer developed a computerized training program that taught participants how to use natural frequency methods to solve problems that required Bayesian reasoning. Their results supported Hoffrage and Gigerenzer's results and showed that when information was presented in

---

[3] Richards J. Heuer, Jr., *Psychology of Intelligence Analysis* (Langley, Virginia: Central Intelligence Agency, 1999), 112.

[4] Ulrich Hoffrage and Gerd Gigerenzer, "Using natural frequencies to improve diagnostic inferences," *Academic Medicine* 73 (1998): 539.

[5] Sedlmeier and Gigerenzer, "Teaching Bayesian Reasoning," 381.

terms of natural frequencies, Bayesian reasoning became easier for people to perform and understand.[6]

This thesis agrees that natural frequencies make it easier for people to perform Bayesian reasoning and builds off of Sedlmeier and Gigerenzer's idea by introducing a paper and pencil method for intelligence analysts that promotes Bayesian reasoning using natural frequencies. The purpose of this study is to determine whether or not the natural frequency tree method introduced in this thesis will lead to better results when compared to Bayesian problems that are worded using natural frequencies. In order to determine this, this thesis will test the following two hypotheses:

Hypothesis One:

> Intelligence-related Bayesian problems worded using natural frequencies will elicit higher Bayesian reasoning amongst intelligence analysts compared to intelligence-related Bayesian problems that are worded using traditional statistical language.

Hypothesis Two:

> A paper and pencil frequency tree method that utilizes natural frequencies can easily be taught to intelligence analysts within 1 ½ hours and elicit a higher level of Bayesian reasoning than Bayesian problems that are only worded using natural frequencies.

---

[6] Ibid, p. 388

In order to answer these hypotheses, three different experiments were performed. The first experiment tested intelligence analysts' performance on Bayesian problems that were worded using natural frequency compared to Bayesian problems worded using traditional statistical language. This experiment was designed to test the first hypothesis by determining if intelligence analysts were able to correctly answer more Bayesian problems when the questions were presented using natural frequencies compared to when they were not worded using natural frequencies. The second experiment built off of the idea of natural frequencies and tested a Bayesian reasoning method using a natural frequency tree format. This second experiment tested the second hypothesis by comparing performance results obtained from using the natural frequency tree method to performance results obtained from wording Bayesian problems using natural frequencies. A third experiment was performed to address some concerns raised about the methodology and results from the second experiment. This experiment was essentially the same as the second experiment, but there were minor changes made to the methodology and instruments used that attempted to correct for the problems.

An easy to use method for Bayesian reasoning that utilizes natural frequencies and is easy for intelligence analysts to use for intelligence-related problems will provide intelligence analysts, both those with a statistical background and those without, with a technique that they can use almost anywhere and one that meets the requirements for analytic techniques laid out in the Intelligence Reform and Terrorism Prevention Act of 2004.

This thesis will be laid out in the following order:

- Chapter 1 of this thesis will provide a quick introduction to the study's purpose, significance, and background.
- Chapter 2 will discuss the available literature on the topic, provide a critical analysis of the different points of view, and detail the rationale for the current study.
- Chapter 3 will describe in detail the methodologies for the three experiments performed in order to determine the answer to the proposed research question.
- Chapter 4 discusses the findings from the three experiments and provides specific conclusions based on these results.
- Chapter 5 will present the overall conclusions from the research done in this study, its implications, and recommendations for future research and application.

CHAPTER 2:

LITERATURE REVIEW

The discussion to use Bayesian reasoning in intelligence analysis is not a recent occurrence. Those in the intelligence field have been writing about its applicability since the early 1970s. However, due to the relatively recent "intelligence failures," especially those revolving around the terrorist attacks of 9/11, discussion to use structured analytical methodologies has been increasing. On December 17, 2004, President George W. Bush signed into law the Intelligence Reform and Terrorism Prevention Act, the most dramatic national intelligence capabilities reform since the signing of the National Security Act of 1947 by then President Harry S. Truman. Based on suggestions from the 9/11 Commission's report and the Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction (The WMD Report), this new law calls for better cooperation throughout the intelligence community and better analytical integrity.[7] Specifically, the Act calls for the use of analytic methodologies, tradecraft, and practices that provide timely and objective intelligence.[8] One method that meets these requirements is Bayesian reasoning.

Bayesian reasoning in general is not a new phenomenon. It has been around since the 18th century when it was first developed by the English clergyman and mathematician Thomas Bayes. Bayes was a non-conformist Minster and a Fellow of the Royal Society. It was not until after his death in 1764 that Bayes' Theorem was published, and his conclusions were not accepted by Pierre-Simon Laplace, the man responsible for soundly

---

[7] United States Government, "Intelligence Reform and Terrorism Prevention Act of 2004."
[8] Ibid.

establishing the theory of mathematical probability[9], until 1781.[10] Today, Bayes is primarily remembered for his work on non-traditional statistical problems.[11]

Simply put, Bayes' Rule is a mathematical formula used to calculate conditional probabilities. It is used to determine the actual chance of an event happening or not happening based on the available information. The idea of how Bayes' Rule can be applied is best illustrated by David Eddy's famous mammography problem. The following problem is an adapted version of Eddy's original mammography problem:

> The probability of breast cancer is 1% for a woman at age forty who participates in routine screening. If a woman has breast cancer, the probability is 80% that she will get a positive mammography. If a woman does not have breast cancer, the probability is 9.69% that she will also get a positive mammography. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?[12]

Most people attempting to answer this question would incorrectly assume that there was between a 70 and 80 percent chance that the woman had breast cancer.[13] The actual answer, however, is much lower. According to Bayes' Rule, there is only a 7.8 percent chance that the woman actually has breast cancer.[14] The reason why the actual answer is so much lower is because Bayes' Rule incorporates the probabilities of both the false positive rate and the hit rate, two items that many people do not take into consideration when attempting to answer this type of problem using their own mathematical calculations.

---

[9] J.J. O'Connor and E.F. Robertson, "Pierre-Simon Laplace," School of Mathematics and Statistics University of St. Andrews, Scotland, http://www-groups.dcs.st-andrews.ac.uk/%7Ehistory/Mathematicians/Laplace.html.
[10] J.J. O'Connor and E.F. Robertson, "Thomas Bayes," School of Mathematics and Statistics University of St. Andrews, Scotland, http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Bayes.html.
[11] Ibid.
[12] Gerd Gigerenzer and Ulrich Hoffrage, "How to Improve Bayesian Reasoning Without Instruction: Frequency Formats," *Psychological Review* 102, no. 4 (1995): 685.
[13] Ibid.
[14] Ibid.

Bayes' Rule is considered to be a scientific method since "it is a tool of statistical inference, used to deduce the probabilities of various hypothetical causes from the observation of a real event."[15] For most people, this rule is not easily understood and proves to be difficult to apply. This is partly because "it forces us [people] to confront and overcome strong biases in our [their] natural way of thinking and … because it is not easy to be specific about exactly where Bayes' Rule will apply, or how it may apply in any particular case."[16]

According to Chris Westbury, author of the paper *Bayes' For Beginners*, "many studies have shown that people of all kinds – even those who are trained in probability theory – tend to be very poor at estimating conditional probabilities. It seems to be … [an] innate incompetence in our species."[17] While some researchers and authors agree with Westbury and argue that human beings are incapable of performing Bayesian reasoning, others argue the opposite.

This literature review will examine both of these opposing views on human beings' abilities to understand Bayes' by discussing the available research on both topics and attempt to identify their strengths and weaknesses. In addition to these topics, this literature review will also address the reasons why Bayes is an excellent analytical solution as well as the body of available writings on the use of Bayesian reasoning in the intelligence field.

Heuristics and Predictions

[15] Nicholas Schweitzer, "Bayesian Analysis for Intelligence: Some Focus on the Middle East," *Studies in Intelligence* 20, no. 2 (1976): 32, https://www.cia.gov/csi/kent_csi/pdf/v20i2a03p.pdf#page=1.
[16] Chris Westbury, "Bayes' For Beginners," (paper written for author's 4th year Psychometrics class, University of Alberta, 2005), 3.
[17] Ibid, 12.

How do we make predictions and judgments when we are faced with uncertainty? Is there a specific process that we all follow? If so, is it systematic and mathematical like Bayesian statistics or is it more of an instinctive approach? These types of questions, revolving around how people think, have been asked by researchers for years. According to Daniel Kahneman and Amos Tversky, two of the best known researchers on this topic:

> In making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or statistical theory of prediction. Instead they rely on a limited number of heuristics which sometimes yield reasonable judgments and sometimes lead to severe and systematic errors.[18]

The two researchers arrived at this conclusion based on their study of category prediction and numerical prediction for their 1973 article entitled "On the Psychology Of Prediction." In category prediction, the prediction is in a nominal form, such as which team will win the Super Bowl. In numerical prediction, the prediction takes on a numerical form, such as what the Dow Jones Industrial Average will be at the end of the day.[19]

To test category prediction, Kahneman and Tversky performed multiple studies, one of which was their famous engineer-lawyer study. This study focused on the impact of the representativeness heuristic on intuitive predictions. Representativeness is one of several heuristics, which are "cognitive mechanisms by which humans make decisions",[20] that Kahneman and Tversky focus on in their studies on heuristics (others include, but are not limited to, availability, causality, and adjustment and anchoring).[21] According to the

[18] Daniel Kahneman and Amos Tversky, "On The Psychology Of Prediction," *Psychological Review* 80, no. 4 (1973): 237.
[19] Ibid.
[20] John M.C. Hutchinson and Gerd Gigerenzer, "Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet," *Behavioural Processes* 69 (2005): 97.
[21] Representativeness, availability, causality, and adjustment and anchoring are not the only three heuristics that can affect judgment. See Appendix A for a complete list of cognitive biases.

representativeness heuristic, people will make predictions and judgments based on how well the information provided represents the outcome being asked. Heuristics, Kahneman and Tversky argue, "reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations."[22]

In their engineer-lawyer study, the authors presented two subject groups with five descriptions. The subjects were told that these descriptions were randomly selected from a group of 100 successful lawyers and engineers. One subject group was told that the engineer-lawyer group consisted of 30 engineers and 70 lawyers while the other subject group was told that the engineer-lawyer group consisted of 30 lawyers and 70 engineers. In both groups, half of the participants were asked to determine the probability the person described was an engineer and the other half of the participants were asked to determine the probability that the person described was a lawyer. The key aspect of this study is that it provided all of the participants with base rate information, which is information regarding the probability that the person is either an engineer or a lawyer. Base rate information is information that describes the initial chances of a hypothesis happening or not happening before any evidence is taken into consideration.

The key to statistical prediction is taking into consideration prior probabilities when evaluating new evidence to make a prediction. Kahneman and Tversky discovered that neither of their subject groups considered the prior probability information they were provided with and allowed the presence of specific evidence to influence their intuitive predictions.[23] Specifically, the representativeness heuristic influenced the subjects' predictions.

---

[22] Amos Tversky and Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science* 185 (1974): 1124.
[23] Kahneman and Tversky, "On The Psychology Of Prediction," 243.

However, when Kahneman and Tversky gave the subjects descriptions without individuating evidence (which leads to the use of the representativeness heuristic) following the five descriptions mentioned above, the subjects applied their prior knowledge regarding the ratio of engineers to lawyers to their predictions.

When studying numerical predictions, the authors found similar results. In one of their studies that examined numerical predictions, Kahneman and Tversky provided subjects with "descriptive information concerning a set of cases."[24] The descriptive information described college freshman. One group was asked to evaluate the quality of each of the descriptions as they pertained to the students and predict the percentage of students whose descriptions indicated that they had a higher academic ability. A second group was asked to predict the students' grade point average at the end of their freshman year as well as his or her percentile ranking. The authors found no significant difference between the two groups.

From their research on both categorical prediction and numerical prediction, Kahneman and Tversky illustrated that "the ordering of outcomes by [their] perceived likelihood coincides with their ordering by representativeness and that intuitive predictions are essentially unaffected by considerations of prior probability and expected predictive accuracy."[25] The authors attribute this to people's difficulty in applying regression, or statistics, and suggest that "a major source of difficulty is that regression effects typically violate the intuition that the predicted outcome should be maximally representative of the input [given] information."[26]

---

[24] Ibid, 244.
[25] Ibid, 238.
[26] Ibid, 250.

Kahneman and Tversky are not alone in their conclusions. In her article "The Base-Rate Fallacy In Probability Judgments," Maya Bar-Hillel also concludes that people fall victim to the base-rate fallacy by having a "tendency to ignore base rates in favor of, *e.g.*, individuating information (when such is available), rather than integrate the two."[27] In her paper, Bar-Hillel provides an answer to why people have a tendency to ignore base-rates by identifying the "cognitive mechanism [that] leads people to ignore base-rate information in problems of Bayesian inference."[28] Unlike Kahneman and Tversky, Bar-Hillel does not believe that representativeness is the reason or that people lack the ability to integrate base-rate information into their judgments.[29] Instead, she claims relevance is the reason. Bar-Hillel believes that "subjects ignore base-rate information … because they feel that it *should* be ignored … because the base rates seem to them *irrelevant* to the judgment they are making."[30]

Under this assumption, Bar-Hillel argues that people are ordering information by the degree they believe it to be relevant to the problem. If people feel that the given individuating evidence is as equally important as the base rate information, then both will contribute to the person's judgment. If people perceive one to be more important than the other, then they will weight it accordingly. According to Bar-Hillel one of the contributing factors to how people decide to order the information is specificity. Unfortunately, in her article, Bar-Hillel falls short of providing a complete answer.

Researchers Zvi Ginosar and Yaacov Trope found similar results to Bar-Hillel with regards to people ordering information when determining probability judgments and

[27] Maya Bar-Hillel, "The Base-Rate Fallacy In Probability Judgments," *Acta Psycholigica* 44, (1980): 211.
[28] Ibid, 216.
[29] Ibid.
[30] Ibid.

also to Kahneman and Tversky regarding the use of individuating information. In their article "The Effects of Base Rates and Individuating Information on Judgments about Another Person," Ginosar and Trope performed an experiment similar to Kahneman and Tversky's engineer-lawyer study. They found that when subjects used individuating information to determine probability judgments regarding a person's category, they clearly exhibited the base-rate fallacy.[31] According to Ginosar and Trope, "presumably, the information presented in such problems activates well-developed schemata or implicit personality theories that commonly serve to relate personality descriptions to other personal characteristics such as profession and fields of study."[32] This conclusion coincides with Kahneman and Tversky's conclusion that people rely on representativeness when making predictions and judgments.

While Ginosar and Trope agreed with Kahneman and Tversky with the conclusion that people will ignore base rates when specific individuating information is provided, they did not agree with Kahneman and Tversky's conclusion that "when worthless specific evidence is given, prior probabilities are ignored."[33] Instead, Ginosar and Trope found that when subjects were given data with inconsistent and unrelated information, subjects incorporated base rate information into their judgments. Ginosar and Trope concluded that:

> While people prefer to base their judgments on schematic inferences from personality descriptions, they nevertheless check the actual content of this singular information for its usefulness for their judgmental purpose …. To the extent that the individuating information does not fulfill this requirement, attention is shifted to alternative, perhaps less preferred

[31] Zvi Ginosar and Yaacov Trope, "The Effects of Base Rates and Individuating Information on Judgments about Another Person," *Journal of Experimental Social Psychology* 16 (1980): 228.
[32] Ibid, 239.
[33] Kahneman and Tversky, "On The Psychology Of Prediction," 242.

> sources of information, namely, the base rate frequencies of outcome categories.[34]

Ginosar and Trope's findings that people do not always ignore the base rate are similar to Bar-Hillel's findings. Like Bar-Hillel, Ginosar and Trope believe that people are capable of evaluating evidence and determining its relevancy. However, unlike Bar-Hillel, who believes that people discriminate information based on specificity, Ginosar and Trope attribute people's evaluation of evidence to a sampling rule, which they believe most people tend to have. Their definition of a sampling rule is people's ability to "properly translate relative frequencies to probabilities."[35] However, according to Ginosar and Trope, if people do not have this sampling rule, they will exhibit the base-rate fallacy regardless of the individuating information's context.

Two other researchers who also determined that people neglect base rates are Richard E. Nisbett and Eugene Borgida. However, instead of studying category membership like Kahneman and Tversky and Ginosar and Trope, Nisbett and Borgida studied people's predictions with regards to behavior. In their paper "Attribution and the Psychology of Prediction," the authors concluded that even when people were given consensus information (i.e. information that described how most people within a particular population behaved), it had no more of an affect on their predictions regarding the behavior of an individual within that population than if they had not received any consensus information.[36] Nisbett and Borgida also found that people were likely to infer a specific type of behavior about a population based only on samples consisting of as few as two people from that population. Their conclusions "thus provide another

---

[34] Ginosar and Trope, "The Effects of Base Rates and Individuating Information," 240.
[35] Ibid.
[36] Richard E. Nisbett and Eugene Borgida, "Attribution and the Psychology of Prediction," *Journal of Personality and Social Psychology* 32, no. 5 (1975): 942.

confirmation of Kahneman and Tversky's … characterization of people as being uninfluenceable by certain types of logically compelling information and overly influenceable by certain types of logically very weak information."[37]

Unlike Ginosar and Trope, who believe that representativeness is responsible for people's intuitive judgments, and Bar-Hillel, who believes relevance is the underlying cause for base-rate fallacy, Icek Ajzen argues that people rely on a causality heuristic when making judgments. In his paper "Intuitive Theories of Events and the Effects of Base-Rate Information on Prediction," Ajzen concludes that:

> When asked to make a prediction, people look for factors that would cause the behavior or event under consideration. Information that provides evidence concerning the presence or absence of such causal factors is therefore likely to influence predictions. Other items of information, even though important by the normative principles of statistical prediction, will tend to be neglected if they have no apparent causal significance. Statistical information is used mainly when no causal information is available.[38]

Ajzen's deduction regarding people's use of a causal heuristic is also supported by Amos Tversky and Daniel Kahneman's conclusion in their paper "Causal Schemas in Judgments Under Uncertainty."

According to Tversky and Kahneman, "base-rate neglect largely depends on whether or not the evidence is given a causal interpretation [by the person making the prediction or judgment]."[39] If the evidence fits into the person's pre-conceived causal schema, then the person will utilize it to make a judgment. On the other hand, if evidence

---

[37] Ibid.

[38] Icek Ajzen, "Intuitive Theories of Events and the Effects of Base-Rate Information on Prediction," *Journal of Personality and Social Psychology* 35, no. 5 (1977): 304.

[39] Amos Tversky and Daniel Kahneman, "Causal Schemas in Judgments Under Uncertainty," in *Progress in Social Psychology* Vol. 1, ed. Martin Fishbein (Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers, 1980), 63.

of equal importance is also provided, such as base-rate information, but it does not fit in the person's causal schema or if the evidence conflicts with it, the person will ignore it.

Tversky and Kahneman showed this through the famous Cab Problem where they told participants that a cab from either the Blue cab company or a cab from the Green cab company was involved in a hit and run accident and that there was also an eye witness that saw the accident. In addition to this information, they also provided the participants with information regarding the percentage of Blue and Green cabs in the city as well as information regarding the eye witness's ability to identify between the different cabs under visibility conditions similar to the time of the accident. Tversky and Kahneman then asked the participants for the probability that the cab involved in the accident was a Blue cab rather than a Green cab.

Tversky and Kahneman concluded that participants ignored the base-rate information and instead used diagnostic information that described the probabilities of the cab being either Blue or Green, because it fit into their causal schema. According to them, when "causally relevant evidence [is present], base-rate data [only] affect predictions when they induce a causal model which (1) explains the base-rate, and (2) applies to the individual case."[40]

In addition to causality and representativeness, the availability and anchoring heuristics are commonly stated as being used frequently by people to make predictions. In his book *Psychology of Intelligence Analysis*, Richards J. Heuer, Jr. claims people use the availability heuristic "whenever they estimate frequency or probability on the basis of how easily they can recall or imagine instances of whatever it is they are trying to

[40] Ibid, 70.

estimate."[41] However, this can lead to incorrect judgments and predictions. People's ability to recall events is directly related to how recent the event actually occurred and whether or not the person was personally involved. These types of factors can cause one to believe that the probability of an event happening is higher than the actual probability of the event occurring. For many people, they do not even realize they are using the availability heuristic.

In addition to availability, Heuer also discusses the anchoring heuristic in his book. The anchoring heuristic occurs when people use some natural starting point, usually based on a previous judgment or belief and then update that belief or judgment based on new evidence. "Typically, … the starting point serves as an anchor … that reduces the amount of adjustment, … [therefore causing] … the final estimate [to stay] … closer to the starting point than it ought to be."[42]

Tversky and Kahneman also agree with Heuer. They too identified the impact of the availability and anchoring heuristic on people's judgments. In their paper "Judgment under Uncertainty: Heuristics and Biases," Tversky and Kahneman discuss different biases that can lead people to use the availability heuristic. According to them, biases due to the retrievability of instances, the effectiveness of a search set, imaginability, and illusory correlation lead people to make incorrect judgments and assessments of probabilities when using the availability heuristic.[43] The retrievability of instances, according to the two authors, can be affected by the familiarity of an event as well as the event's salience and recentness. In regards to the effectiveness of a search set, Tversky and Kahneman are referring to people's abilities to extract information from a specific set

[41] Heuer, Jr., *Psychology of Intelligence Analysis*, 147.
[42] Ibid, 150.
[43] Tversky and Kahneman, "Judgment under Uncertainty," 1127-8.

of data, for example, whether or not a word with the letter *r* as the third letter is more common than a word with the letter *r* as the first letter.[44] Biases of imaginability refer to one's ability to assess the frequency of an event that is not stored in memory. To do this the person has to generate the event's frequency based on his or hers ability to construct events. Finally, the illusory correlation bias occurs when "the judgment of how frequently two events co-occur … [is] based on the strength of the associative bond between them. When the association is strong, one is likely to conclude that the events have been frequently paired."

In regards to anchoring, Tversky and Kahneman discuss different situations upon which the anchoring heuristic can have a negative effect. One such situation they discuss is insufficient adjustment. To illustrate their point, the authors use two numerical examples:

$$8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$$

$$1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$$

When their subjects were given these two expressions and asked to rapidly answer them without a calculator, they underestimated both expressions; however, there was a drastic difference between the two estimations. The estimation for the problem with descending numbers was significantly higher than the estimation for the problem with ascending numbers. This occurred because the subjects performed the first few steps of the problem and then estimated the actual product by using adjustments. Since the first problem started off with higher numbers, the number that the subjects used as an anchor was higher than the number the subjects who received the second problem used as an anchor. As indicated earlier by Heuer, people's adjustments tend to stay close to their selected

[44] Ibid, 1127.

anchor; and therefore, due to insufficient adjustment, the subjects extremely underestimated the answers to the two problems.

While there is a large amount of research on the use of heuristics in people's ability to make predictions and judgments when faced with uncertainty, Heuer's book is the only one that addresses the impact of heuristics on intelligence analysts' abilities to make predictions. Even though no other author directly applies heuristics to the intelligence field, it is safe to say that all of their research can easily be applied to intelligence analysts. Intelligence analysts are like any other human being; even "experienced researchers are … prone to the same biases [that everyone else is] – when they think intuitively."[45]

While the notable literature described above about how people make decisions is compelling, there are a few weaknesses in these studies. Kahneman and Tversky are inarguably the leaders in the study of judgment and heuristics, and consequently, many researchers cite their work and attempt to replicate their studies. This is definitely essential, but it also leads to very narrow results in terms of applicability. The majority of the studies on heuristics and judgment focus on category membership. While this is an important aspect of decision making, not all decisions pertain to which category a person belongs. The available research lacks insights regarding the processing of more complex judgments and predictions under uncertainty.

Another weakness of the available literature is that it focuses on a handful of heuristics. The literature previously discussed highlights the most commonly cited reasons why people exhibit the base-rate fallacy and should not be considered the only

---

[45] Ibid, 1130.

explanations. There are several other cognitive heuristics that influence peoples' abilities to make predictions (see Appendix A for a list of additional cognitive heuristics).

In addition to the weakness of heuristics, the studies also focus on the inability of people to take into consideration base rates when making predictions. While all of the studies state that base rate information is the most important factor in order to make a correct prediction or judgment based on the evidence, they offer no solution to this problem.

The next section of this literature review will discuss one possibility that physical scientists argue is the solution to this problem. According to them, the use of Bayesian reasoning will help reduce the use of heuristics in decision making and increase the use of base rates.

Bayesian Reasoning

In recent years, Bayesian reasoning "has experienced a strong revival amongst statisticians and philosophers … due, in part, to its intrinsic plausibility and also to the weaknesses which have gradually been exposed in the standard methodologies."[46] In their book *Scientific Reasoning: The Bayesian Approach*, authors Colin Howson and Peter Urbach show that the use of traditional statistical methods has several flaws and that using a Bayesian approach can help avoid some of these difficulties. One of the benefits of using a Bayesian approach over other more traditional forms of statistical inference is that Bayes does not discriminate against different hypotheses. Under the Bayesian approach, all hypotheses, whether they be deterministic or statistical, are treated

---

[46] Colin Howson and Peter Urbach, *Scientific Reasoning: The Bayesian Approach* (La Salle, Illinois: Open Court Publishing Company, 1989), 10.

the same. In addition, using Bayes allows "one to associate different degrees of confidence with different values and ranges of values, as classical statisticians sought to do, but it does this in a simple and straightforward way …."[47] While Bayes does have a subjective approach to determining initial probabilities, the authors argue that because of the way that Bayes requires evidence to be individually evaluated, those with opposing views will "normally approach a common view as the evidence accumulates,"[48] therefore indicating that Bayes reduces the affect of heuristics and initial biases. According to the authors, Bayes' processes for inductive reasoning are both impartial and objective.

Howson and Urbach also provide a rebuttal to Kahneman and Tversky's argument that "in his evaluation of evidence, man is … not Bayesian at all."[49] According to Howson and Urbach, Kahneman and Tversky do not provide any justification for their conclusion that people do not use the precepts of Bayes when processing data. They argue that:

> All the cases which are supposed to show this depend on the assumption of the subjects' tacit acceptance of some type or other of [a] probability model which the testers [Kahneman and Tversky] think appropriate. If the subjects had chosen some different model, for whatever reason, then the conclusions to be drawn may be entirely different.[50]

The benefits of using Bayes can also be found in Gudmund R. Iversen's book *Bayesian Statistical Inference*. In his book, he describes some of the strengths of Bayesian inference as well as an argument as to why people should use Bayesian inference over classical statistical inference. The main advantage Iversen discusses is that Bayes can be used by those who have a very limited understanding of statistics.

---

[47] Ibid, 253.
[48] Ibid, 254.
[49] Daniel Kahneman and Amos Tversky, "Subjective Probability: a Judgment of Representativeness," *Cognitive Psychology* 3, no 3 (1972): 450.
[50] Howson and Urbach, *Scientific Reasoning*, 293.

Very few people take more than one statistics course in their undergraduate education. Additionally, he states, it is rare that statistical inference will be covered and when it is, most people have a very hard time understanding concepts such as significance levels and confidence intervals. According to Iversen, "Bayesian inference is more natural and the conclusions [are] much easier to understand."[51] The reason why Bayesian inference is more natural is because "it is more closely geared to the research process itself than classical inference."[52]

In the research process, one starts with an initial uncertainty or hypothesis. He or she then collects data to reduce the uncertainty and make a better decision regarding the hypothesis. The same type of approach is used in Bayesian inference with one important addition. Bayes provides a numerical component by computing the probabilities of the data and hypotheses. The benefit of using Bayes over classical inference is that it expresses the probabilities of items that are truly uncertain, unlike classical inference that computes the probability of observed data or something that is already known to be true.[53]

In his article "An Intuitive Explanation of Bayesian Reasoning," Eliezer Yudkowsky builds off of the idea of Bayes following the research process. In his discussion, he highlights the Bayesian method's use of evidence to evaluate hypotheses. According to Yudkowsky, Bayes' theorem is capable of distinguishing between weak and strong evidence.[54] In other words, Bayes can indicate what is evidence and what is not evidence and also evaluate the evidence's strength. In addition, he claims that Bayes' not only indicates when to revise the hypotheses, but also how to revise them.

---

[51] Gudmund R. Iversen, *Bayesian Statistical Inference*, (Beverly Hills: Sage Publications, Inc., 1984), 75.
[52] Ibid, 76.
[53] Ibid.
[54] Eliezer Yudkowski, "An Intuitive Explanation of Bayesian Reasoning," http://yudkowsky.net/bayes/bayes.html.

Another author who believes Bayes makes understanding probabilities and statistical decision theory easier is Peter M. Lee, the author of *Bayesian Statistics: An Introduction*. In his book, he provides a very good example to illustrate the difficulty of understanding the logic of classical statistical inference. Under classical statistical inference, a 95 percent confidence interval indicates that if similar procedures regarding an unknown condition were carried out "time after time then the unknown conditions would lie in the confidence intervals … constructed 95 … [percent] of the time."[55] To anyone who does not have a significant amount of training in statistics, this means nothing. Most likely if a non-mathematical person was told that there was a 95 percent confidence interval, he or she would assume that there was a 95 percent probability of the event occurring. However, as the earlier statement indicates, this is not true. Under Bayes, however, this is exactly what a 95 percent confidence interval means, thus indicating that the conclusions derived from using Bayesian inference are much easier to understand.

David Malakoff, author of the article "Bayes Offers a 'New Way' to Make Sense of Numbers," also agrees with Lee and Iverson. He too agrees that Bayes produces answers easier for non-statisticians to understand. Bayes theorem "is mathematics on top of common sense."[56] In his article, Malakoff quotes Steven Goodman, a physician and biostatistician from Johns Hopkins University, who states that, "Bayesian computations give you a straightforward answer you can understand and use …. It says there is an X% percent probability that your hypothesis is true,"[57] as opposed to the example earlier regarding confidence intervals. According to Malakoff, another benefit of using Bayes is

---

[55] Peter M. Lee, *Bayesian Statistics: An Introduction*, (London: Arnold Publishers, 1997), ix.
[56] Actual quote from Kathryn Blackmond Laskey of George Mason University but taken from David Malakoff, "Bayes Offers a 'New Way' to Make Sense of Numbers," *Science* 286 (1999): 1460.
[57] David Malakoff, "Bayes Offers a 'New Way' to Make Sense of Numbers," *Science* 286 (1999): 1464.

that Bayes allows people to factor their previous knowledge and expertise into their computations, which is something that traditional statistical methods disapprove. Due to this, it forces people to be explicit about their biases, which other more traditional methods tend to obscure.[58]

As the literature described above indicates, using a Bayesian reasoning method forces people to take base rates into consideration. Bayes does this by requiring people to use their available knowledge regarding the hypothetical events to assign initial probabilities to the hypotheses. In addition to requiring people to utilize base rates, Bayes also helps reduce biases by causing people to individually evaluate each piece of evidence against the different hypotheses. As mentioned earlier by both Malakoff and Howson and Urbach, "enough evidence will lead people who start with dramatically different priors [likely due to their use of heuristics] to essentially the same posterior answer."[59] The use of Bayesian reasoning thus provides people with a systematic reasoning method that yields results which are easier for someone without at strong mathematical background to understand.

Bayesian Reasoning and the Intelligence Community

The applicability of Bayes as an analytic methodology for analysts within the intelligence community is not a recent or new occurrence. While lately it has picked up some speed due to the "intelligence failures" of 9/11 and the incorrect claim of weapons of mass destruction (WMD) in Iraq, analysts and those within the intelligence community were writing about the use of Bayes' method in the early 1970s. In 1972, Jack Zlotnick,

[58] Ibid.
[59] Ibid, 1461.

a former analyst in the Central Intelligence Agency (CIA), wrote a paper entitled "Bayes Theorem for Intelligence Analysis." Zlotnick's paper was written at the time when the CIA was first studying the applicability of a Bayesian method for intelligence analysis. In his paper, Zlotnick explains the reason why the intelligence community needs a quantitative methodology like Bayes for analysis.

According to Zlotnick, all intelligence analysts most likely base their estimates on incomplete evidence; therefore "the very best that intelligence [analysts] can do is to make the most of the evidence without making more of the evidence than it deserves."[60] One way of accomplishing this is to use probabilities, specifically Bayes' Theorem.[61] Bayes' Theorem, according to Zlotnick complements other analytical approaches well, especially relating to strategic warning. In the intelligence field, this is one of the sole reasons why decision makers want estimative analysis. "Strategic warning analysis focuses primarily on just the problem that Bayes' Theorem addresses – the odds favoring one hypothesis (an imminent attack) over another hypothesis (no imminent attack).[62]

According to Zlotnick, there are three distinguishing benefits of using a Bayesian method for intelligence analysis. The first benefit is that it forces analysts to quantify judgments to which they usually do not assign numerical values. The second benefit of Bayes is that it allows the analyst to evaluate the evidence against the hypotheses and not take the evidence as a given upon which to draw conclusions. The third feature of using a Bayesian method is that it allows analysts to make judgments about bits and pieces of the evidence and not "sum up the evidence as he [or she] would have to do if he [or she] had to judge its meaning for final conclusions. The [overall] mathematics [from using

[60] Jack Zlotnick, "Bayes' Theorem For Intelligence Analysis," *Studies in Intelligence* 16, no. 2 (1972): 43, https://www.cia.gov/csi/kent_csi/pdf/v16i2a03p.pdf#page=1.
[61] Ibid.
[62] Ibid, 45.

the Bayesian method] does the summing up, telling the analyst in effect: 'If these are your readings of the individual items of evidence, then this it the conclusion that follows.'"[63] Zlotnick feels that Bayes has several advantages to offer and is a step in the right direction. However, he feels that Bayes is not a complete solution, but only a "path to improvement."[64]

As Zlotnick mentions in his paper, there is an argument for the use of Bayes' over traditional analytical methods in intelligence warning. In his paper, "The Sino-Soviet Border Dispute: A Comparison of the Conventional and Bayesian Methods for Intelligence Warning," Charles E. Fisk performed an experiment to test this assertion. Fisk discovered that the Bayesian approach to intelligence warning provided a type of "accounting system" for intelligence analysis. According to Fisk, "if such a system were implemented for other questions [questions outside of the Sino-Soviet border dispute questions], a significant class of disagreements among analysts might be resolved."[65] Using a Bayesian method allows analysts to see how other analysts arrived at their conclusions by showing how each analyst individually evaluated and weighed the evidence. In the end, Fisk concluded that "an improved system of accounting for analytical judgments,"[66] such as Bayes, was needed in the intelligence community.

In his article "Bayesian Analysis for Intelligence: Some Focus on the Middle East," Nicholas Schweitzer also looks at the applicability of Bayesian inference in the intelligence field. In his study, he uses Bayes to perform political analysis. Schweitzer also indicates that his study is the first one to apply Bayes in a non-historical context. At

---

[63] Ibid.
[64] Ibid, 52.
[65] Charles E. Fisk, "The Sino-Soviet Border Dispute: A Comparison of the Conventional and Bayesian Methods for Intelligence Warning," *Studies in Intelligence* 16, no. 2 (1972): 61.
[66] Ibid, 62.

the time he wrote his paper, the CIA had been investigating the utility of Bayesian inference for many years and had accepted its applicability based on earlier experiments that applied “Bayes to the analysis of historical intelligence situations.”[67]

In his study, Schweitzer used a group of analysts. He had each analyst assign a set of probabilities to four hypotheses estimated to occur within 30 days regarding situations in the Middle East. It was up to the analyst to determine the probabilities for each hypothesis based on his or her understanding of the situation at the time. The probabilities for the four hypotheses had to add up to one or 100 percent. The values that the analysts initially assigned were imprecise, but they provided an excellent starting point for comparison and analysis. Once these initial percentages were chosen, the analysts were then allowed to assess both open source and classified evidence and change their estimates accordingly using Bayes’ rule. Schweitzer argues that using the simplified version of Bayes’ rule is quite easy;[68] therefore indicating analysts should not have a problem using Bayesian inference.

According to Schweitzer, “the Bayesian method upon completion results in an archive of evidence, evaluations, and predictions which lend themselves to various forms of evaluation.”[69] Additionally, Schweitzer argues that Bayes’ “ability to portray the results of the analysis graphically … [is] one of the strongest arguments for using it …. [Bayes] conveys much information at a glance, and seems to represent an advance in communication over traditional methods of reporting, especially in illustrating trends far more concisely and vividly than do words”[70] Schweitzer also concluded that the use of

---

[67] Schweitzer, “Bayesian Analysis for Intelligence, 33-34.
[68] Ibid, 32.
[69] Ibid, 39.
[70] Ibid, 35.

Bayes in intelligence analysis indicated changes in the hypotheses earlier than an analyst's intuitive judgment would likely indicate.[71]

Like Zlotnick, Schweitzer also points out specific benefits of using a Bayesian method in intelligence analysis. The first benefit is that Bayes forces analysts to assess all pieces of evidence in a systematic way thus eliminating any biases resulting from recentness or visibility. The second benefit is that a Bayesian procedure is transparent and therefore can be reproduced by other analysts who may disagree with the final estimate. A third benefit is that the way in which Bayes causes questions to be formulated, "forces the analyst to consider alternative explanations of the facts he [or she] sees …."[72] A fourth benefit is that it forces numerical assignments, which are easier to interpret than words such as "likely" or "unlikely." Finally, according to Schweitzer, Bayes "has been shown to be less conservative than analysts' informal opinions, and to drive the probabilities away from 50-50 faster and farther than the analysts' overall subjective judgments do."[73] While some of Schweitzer's benefits are different, there are also some that are similar to Zlotnick's benefits of using a Bayesian method.

Schweitzer also agrees with Zlotnick that Bayes is not the absolute solution for intelligence analysis. According to him, Bayes is thought by many people to be very thought-provoking and an advance in communications compared to traditional analysis;[74] however, since Bayes can only be used in certain situations, and there are some limitations to the method. These include subjective determination of relevant evidence and source reliability problems. Therefore, it can only be a compliment to traditional analysis.

[71] Ibid, 40.
[72] Ibid.
[73] Ibid.
[74] Ibid, 41.

It appears from the literature described up to this point that the majority of study done on Bayesian analysis by the intelligence community itself was performed in the 1970s. In 2005, however, Jessica McLaughlin once again investigated Bayesian analysis and its applicability in the intelligence community. For her thesis, McLaughlin discussed the use of a Bayesian updating model and performed a case study using the technique on Iraq's nuclear weapons program. In her evaluation of the Bayesian updating model, she concluded that "one of the biggest advantages of using a Bayesian framework for intelligence analysis is … [its] inherent transparency; it is clear to the observer how the analyst came to the decision that her or she did and what judgments were made in that process."[75] In addition, McLaughlin states that Bayesian analysis helps ensure that events, both supporting and contradictory to the hypotheses, are not overlooked. She also, like many of the other authors, agrees that Bayes is not the ultimate solution, but an excellent supplementary technique to help intelligence analysts provide decision makers with the most complete analysis possible.

The available literature about the use of Bayes in the intelligence community is very limited. This is surprising because it is "often recommended to analysts."[76] According to Michael Schrage, Senior Advisor to MIT's Security Studies Program:

> It's time to require national security analysts to assign numerical probabilities to their professional estimates and assessments as both a matter of rigor and record. Policymakers can't weigh the risks associated with their decisions if they can't see how confident analysts are in the evidence and conclusions used to justify those decisions.[77]

---

[75] Jessica McLaughlin, "A Bayesian Updating Model for Intelligence Analysis: A Case Study of Iraq's Nuclear Weapons Program" (Thesis, Stanford University, 2005), 52.
[76] Kristan J. Wheaton, Jennifer Lee, and Hemangini Deshmukh, "Bayesian Statistics In The Real World Of Intelligence Analysis: Lessons Learned" (paper presented at the International Studies Association's 2007 Annual Convention, Chicago, Illinois, February 28 – March 3, 2007).
[77] Michael Schrage, "What Percent is 'Slam Dunk'?," *Washington Post*, February 20, 2005, Page B01, http://www.washingtonpost.com/wp-dyn/articles/A37115-2005Feb19.html.

A solution to this, according to Schrage is Bayes. He describes Bayes as a "powerful tool to weigh new evidence."[78] Supporting this claim is Bruce Blair, President of the World Security Institute and Board of Directors member of the Center for Defense Information. According to him, Bayes is a "rigorous approach … [that] provides an account of how the required judgment, or interpretation [for estimating the probability of an event], might be made in a disciplined, responsible manner."[79] In addition, Blair argues that "Bayesian analysis goes a long way toward explaining the seemingly flawed risk assessments made by the intelligence community leading up to 9/11 and the Iraq war."[80]

While the intelligence community is familiar with Bayes and its benefits to the technique, as indicated by the previous literature, Bayes still has not "seeped into the national security community's analytical mainstream."[81] The 9/11 Commission Report, the WMD Report, and the Intelligence Reform and Terrorism Prevention Act of 2004 have all called for "improvement in intelligence analysis and the tools and methods that analysts use to create their estimates."[82] Why then is Bayes not being used in the intelligence community? The next section attempts to answer this question.

<u>The Difficulty of Using Bayesian Reasoning in the Intelligence Community</u>

If there are so many benefits of using Bayesian reasoning, there must be some pretty good reasons why the intelligence community is not applying this analytical technique. In addition to the argument that people's heuristics and biases prevent them from using Bayesian reasoning, the most common reason cited is that it is too difficult to

---

[78] Ibid.
[79] Bruce Blair, "The Logic of Intelligence Failure," Center for Defense Information, March 9, 2004, http://www.cdi.org/blair/logic.cfm.
[80] Schrage, "What Percent is 'Slam Dunk'?."
[81] Ibid.
[82] Wheaton, Lee, and Deshmukh, "Bayesian Statistics In The Real World."

use. Bayes' Theorem is considered advanced statistics and is not usually covered in undergraduate or graduate statistics courses (barring those studying in the mathematical field). The uneasiness regarding the use of Bayes can easily be seen by looking at the Bayes' Theorem[83] formula itself:

$$P(H|D) = \frac{P(H)*P(D|H)}{(P(H)*P(D|H))+(P(NH)*P(D|NH))}$$

Where:

P(H|D) is the probability that the hypothesis is true given the data.

P(H) is the probability of the hypothesis being true.

P(D|H) is the probability of the data given that the hypothesis is true (hit rate).

P(NH) is the probability that the hypothesis is not true.

P(D|NH) is the probability of the data given that the hypothesis is not true (false positive).

For most analysts, this formula makes no sense at all. In addition, "Bayesian math points to a fairly slow learning curve."[84] Even if he or she wanted to learn more about Bayes,

[83] Sedlmeier and Gigerenzer, "Teaching Bayesian Reasoning," 382-4.
[84] Blair, "The Logic of Intelligence Failure."

the majority of literature available on the topic is intended for those with a strong mathematical background who have at least learned advanced calculus, if not more. Due to this mathematical complexity, analysts tend to shy away from the use of Bayes.

In addition to this reason, there are also other reasons within the intelligence community not related to Bayes' mathematical complexity that have prevented it from being used. In MSgt Robert D. Folker, Jr.'s paper "Intelligence Analysis in Theater Joint Intelligence Centers: An Experiment in Applying Structured Methods," he states that the complexity of some scientific methodologies, like Bayes, and the intelligence community's lack of expertise in exploiting them is what prevents them from being used by intelligence analysts. He says:

> A major issue common to all the methods is customer acceptance. Due to the highly mathematical nature of [Bayesian Decision Analysis], many users will feel uneasy trusting the resulting assessments. This will only be overcome through proper training of the analysts using [Bayesian Analysis] and repeated exposure to Bayes on the part of decision makers.[85]

Another reason Folker states for intelligence analysts not to using Bayes as well as other scientific methodologies is that they are "too cumbersome."[86] Folker further elaborates by saying that analysts feel that the increased used of these methodologies leads to more accountability, something that analysts are hesitant to take upon themselves. Most analysts prefer instead a subjective approach that is more intuitive and non-structured.[87]

In his ethnographic study of the intelligence culture, Dr. Rob Johnston found some additional reasons why intelligence analysts were unlikely to use techniques such as Bayes. For part of his study, he asked analysts to describe their analytic process.

---

[85] Quote actually from Captain David Lawrence Graves' (USAF) MSSI *Bayesian Analysis Methods for Threat Prediction* (July 1993) but was cited in MSgt Robert D. Folker, Jr. (USAF), "Intelligence Analysis in Theater Joint Intelligence Centers: An Experiment in Applying Structured Methods" (Thesis, Joint Military Intelligence College, 2000), 8.
[86] Folker, Jr., "Intelligence Analysis," 14.
[87] Ibid, 2.

According to one analyst he interviewed, analysts have "Bayesian tools, simulations, [and] all kinds of [other] advanced methods,"[88] but they do not have any time to use them. Another analyst states, "I don't have time to worry about formal analytic methods. I've got my own system. It's more intuitive and a lot faster."[89]

Another reason that analysts tend to shy away from approaches such as Bayes is that they prefer to use words instead of numbers to describe their analytic confidence levels. Additionally, some analysts feel that it is difficult to apply numerical analysis to intelligence judgments.[90] It is the same for decision makers. According to a senior CIA officer who has served for more than 20 years, "most consumers of intelligence aren't particularly sophisticated when it comes to probabilistic analysis. [Additionally], … [they] prefer briefings that don't center on numerical calculation."[91]

Unfortunately, there is no immediate solution that will remedy the hesitation amongst analysts and decision makers to use more scientific and numerical methods of analysis for the creation of intelligence estimates. This is a change that both must decide to do on their own. Fortunately, there is a solution for the mathematical complexity of Bayes. According to Gerd Gigerenzer and many of his colleagues, if Bayesian problems were worded in terms of natural frequencies, then it would be easier for people to perform Bayesian reasoning. The next section of this literature review will discuss their research on the applicability of natural frequencies in Bayesian reasoning.

Bayesian Reasoning and Natural Frequencies

[88] Rob Johnston, *Analytic Culture in the U.S. Intelligence Community: An Ethnographic Study* (Washington, DC: Center for the Study of Intelligence, 2005): 15.
[89] Ibid.
[90] Heuer, Jr., *Psychology of Intelligence Analysis*, 152.
[91] Schrage, "What Percent is 'Slam Dunk'?."

Natural frequencies are a way of representing statistical information in a way that people without a strong mathematical background can understand. For example, instead of saying 10 percent, one could use natural frequencies and explain the same percentage by saying 10 out of 100. What makes natural frequencies easier to understand is the fact that they provide population information; that is, they are not normalized with respect to base rates. According to Leda Cosmides and John Tooby from the Center for Evolutionary Psychology at the University of California, Santa Barbara, using natural frequencies is advantageous because they preserve "important information that would be lost by conversion to a single-event probability."[92] In addition, "natural frequencies correspond to the way humans have encountered statistical information during most of their history."[93] According to Gerd Gigerenzer, the most well-known researcher on the topic of natural frequencies, and his collaborators, the capacity for inductive inference in the human mind has been built up through evolution to use natural frequencies.[94] This is because throughout evolution, humans encountered frequencies of actual events occurring (i.e. successfully catching a rabbit two out of ten attempts). Cosmides and Tooby also agree with Gigerenzer that from an evolutionary and functional point of view, it makes sense that the human mind would be able to use natural frequencies.[95]

For their paper "How to Improve Bayesian Reasoning Without Instruction: Frequency Formats," Gerd Gigerenzer and Ulrich Hoffrage provide both a theoretical framework that discusses the reasons why natural frequencies improve Bayesian

[92] Leda Cosmides and John Tooby, "Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty," *Cognition* 58 (1996): 16.
[93] Gerd Gigerenzer and Adrian Edwards, "Simple tools for understanding risks: from innumeracy to insight," *BMJ* 327 (2003): 742.
[94] Sedlmeier and Gigerenzer, "Teaching Bayesian Reasoning," 381.
[95] Cosmides and Tooby, "Are humans good intuitive statisticians," 14.

reasoning and studies to support their claim.[96] According to the authors, the studies done by Kahneman and Tversky, Bar-Hillel, and other heuristic supporters are invalid because they are asking mathematical questions that use algorithms different than the cognitive algorithms in the human mind.[97] Gigerenzer and Hoffrage believe that if Bayesian problems are worded using natural frequencies, people will be able to perform Bayesian reasoning to answer the problem because the natural frequencies will easily work with the cognitive algorithms the human mind uses to solve these types of problems.[98] This point is best summed up and supported by Sedlmeier and Gigerenzer:

> The general point here is that Bayesian algorithms are dependent on the information format. Note that the two information formats – probability and frequency – are mathematically equivalent, and so are the two equations; but the Bayesian algorithms are not computationally and psychologically equivalent.[99]

One experiment Gigerenzer and Hoffrage performed to confirm their hypothesis involved testing different problem formats. In this experiment, Gigerenzer and Hoffrage provided participants with questions using both standard probability formats and frequency formats. They found that when they used the standard probability formats, only 16 percent of the participants used Bayesian algorithms.[100] However, when they used the frequency formats, they found that between 46 and 50 percent of the participants used Bayesian algorithms.[101] The results from this experiment show that "Frequency formats, in contrast to probability formats, "invite" Bayesian algorithms …."[102]

[96] Gigerenzer and Hoffrage, "How to Improve Bayesian Reasoning," 685.
[97] Ibid, 684.
[98] Ibid, 685.
[99] Sedlmeier and Gigerenzer, "Teaching Bayesian Reasoning," 382.
[100] Gigerenzer and Hoffrage, "How to Improve Bayesian Reasoning," 693.
[101] Ibid.
[102] Ibid, 695.

According to Gigerenzer and Hoffrage, "Frequency representations can help people "see" the answer …."[103]

In another paper by Hoffrage and Gigerenzer entitled "Using natural frequencies to improve diagnostic inferences," the authors tested whether or not physician's diagnostic inferences could be improved using natural frequencies instead of probabilities.[104] They tested 48 physicians in Munich and Düsseldorf with an average of 14 years professional experience. Hoffrage and Gigerenzer provided the participants with booklets containing four problems. Two of these contained information presented in probabilities and two with information presented using natural frequencies. The formats and order of the problems varied amongst the participants. Once again, Hoffrage and Gigerenzer found that the use of natural frequencies improved Bayesian reasoning performance. They found that the physicians were only able to reason using Bayes' rule 10 percent of the time when the information was presented in probabilities. However, when the information was presented using natural frequencies, they found that the physicians were able to reason using Bayes' rule 46 percent of the time.[105] According to Hoffrage and Gigerenzer, "The physicians spent an average of about 25% more time on the problems involving probabilities than they did on those involving natural frequencies, indicating that they found them more difficult to solve."[106]

In their research, Leda Cosmides and John Tooby also found evidence to support the finding that humans are better able to perform inductive reasoning using natural frequencies. In their paper, "Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty,"

[103] Ibid, 701.
[104] Hoffrage and Gigerenzer, "Using Natural Frequencies," 538-540.
[105] Ibid, 539.
[106] Ibid.

Cosmides and Tooby used a famous problem in the "heuristics and biases" literature and showed that when the problem was worded using natural frequencies, people were better able to perform Bayesian reasoning.[107]

The famous "heuristics and biases" problem that Cosmides and Tooby used was a medical diagnosis problem performed by W. Casscells et al.[108] The medical diagnosis problem provided the participants with information regarding the prevalence of a disease and the false positive rate. Casscells et al. asked the participants to determine the likelihood of a person actually having the disease if they tested positive for it. Only 18 percent of their participants answered correctly and Casscells et al. concluded that their participants violated "Bayes' theorem by ignoring the base rate."[109]

Instead of using the traditional statistical language that Casscells et al. used, Cosmides and Tooby converted the probabilities to natural frequencies and unlike Hoffrage and Gigerenzer, Cosmides and Tooby did not use physicians for their medical diagnosis problem. They instead used undergraduate students for their experiment. The two researchers found that when Casscells et al.'s medical diagnosis problem was converted to natural frequencies, they were able to elicit answers that used Bayesian reasoning from the majority of the participants (76 percent) and that "base-rate neglect virtually disappeared."[110]

All of the previous literature examined the effects of different formats on people's abilities to perform Bayesian reasoning without any instruction. The following literature

---

[107] Cosmides and Tooby, "Are humans good intuitive statisticians," 1.
[108] W. Casscells, A. Schoenberger, and T.B. Graboys, "Interpretation by physicians of clinical laboratory results," *New England Journal of Medicine* 299 (1978): 999-1001.
[109] Cosmides and Tooby, "Are humans good intuitive statisticians," 21-22.
[110] Ibid, 27.

looks at ways to teach Bayesian reasoning to those without a statistical background or a strong familiarity with Bayes in general.

In their paper "Teaching Bayesian Reasoning in Less Than Two Hours," Peter Sedlmeier and Gerd Gigerenzer present and test a new method for teaching Bayesian reasoning.[111] According to the authors, previous studies up to this point have reported little success in regards to teaching this analytical technique. To teach Bayes, Sedlmeier and Gigerenzer used a "computerized tutorial program to train people to construct frequency representations … rather than to insert probabilities into Bayes's rule …."[112]

Sedlmeier and Gigerenzer build off of the previous research already discussed that indicates Bayesian computations are simpler to perform when the information is presented in natural frequencies compared to probabilities and percentages. The difference in their teaching method compared to other teaching methods is that they focus having people construct frequency representations as opposed to applying given rules.

Sedlmeier and Gigerenzer's teaching program had three separate tutorials, which were used on three different subject groups. They also had a control group that received no training. The first tutorial taught rule training. This tutorial showed participants how to insert the information from the problem into Bayes' rule. The other two tutorials explained to participants how to construct either a frequency tree representation or a frequency grid representation.

The frequency tree representation had participants construct a reference class (which indicated the total number of observations) that was broken down into four subclasses. Each subclass represented a frequency that was either given in the problem

[111] Sedlmeier and Gigerenzer, "Teaching Bayesian Reasoning," 380.
[112] Ibid.

or calculated from the information provided in the problem. The computer allowed the participants to choose the reference class size and then helped them construct the subclasses of the tree and performed any required math. Along the way, the computer explained to the participants how the numbers were derived.

In the frequency grid representation, each square of the grid represents one case. The participants started with an empty grid that represented all of the possible cases and then shaded squares depending on the information given in the problem. The computer allowed the participants to choose their own grid sizes and explained to them how to shade the squares.

All three tutorials used the same example problems, which were all in the probability format. After the guided tutorials, the participants were then asked to solve eight additional problems on their own, which utilized problems in both frequency and probabilistic formats. During these problems, the participants received immediate feedback from the computer. Since the participants had to correctly solve a problem before moving on to the next problem, the computer provided assistance, if the participant wanted it, to help him or her correctly step through the problem if he or she was having difficulty or made a mistake. The participant was always given the choice to try and correct the error on his or her own or to have the computer correct it for him or her.

In addition to the testing immediately following the training session, Sedlmeier and Gigerenzer tested the participants on their learning a week later as well as 1-3 months after the training session. They then used the results from all three testing sessions to determine the effectiveness of their training program. They used the immediate testing to test for an immediate training effect, the second test (a week later) for generalization and transfer effects, and the third test (1-3 months later) for stability of learning over time.

Sedlmeier and Gigerenzer found that there were significant immediate learning effects from the training sessions. For the rule based group, their performance increased from being able to use Bayesian reasoning correctly on zero percent of the problems (prior to training) to 60 percent following training.[113] For the frequency groups, their average baseline mean prior to training was 10 percent, but after training, their performance increased to 75 percent and 90 percent.[114] Sedlmeier and Gigerenzer also found that all three training programs lead to high levels of transfer. The participants' average performance on the new problems a week later was almost as good as their performance on the problems immediately following training.[115] However, when Sedlmeier and Gigerenzer looked at stability, they found that performance in the rule base group dropped significantly but performance in the frequency representation groups remained about the same, with a little higher performance found in the frequency grid group.

Due to high attrition rates in the study described above, Sedlmeier and Gigerenzer ran a second study to see if they could replicate the results in the absence of the high attrition rates. In this experiment they found similar results to their earlier experiment.

From this study, Sedlmeier and Gigerenzer hypothesized that the use of graphical representations also made an impact on retention. They tested this hypothesis in a second study that used a probability tree and a frequency tree to perform the Bayesian reasoning. Sedlmeier and Gigerenzer found that students who used the frequency tree representation

[113] Ibid, 388.
[114] Ibid.
[115] Ibid.

were able to retain what they had learned longer than the students who used the probability trees,[116] thus showing the superiority of both natural frequencies and graphical representation in solving Bayesian equations.

Two other researchers support Sedlmeier and Gigerenzer's findings. Laura Martignon and Christoph Wassner from the Max-Planck-Institute for Human Development in Germany also studied how to teach statistical decision making using natural frequencies. In their paper "Teaching Decision Making and Statistical Thinking with Natural Frequencies," Martignon and Wassner also used a computerized training program to teach students how to perform Bayesian reasoning using representations.[117] They then measured the effects of the representation training by using subsequent tests. Like Sedlmeier and Gigerenzer, Martignon and Wassner found that students who used representations that utilized frequencies performed much better than students who used representations that involved probabilities.

All of the literature on the use of natural frequencies indicates that "base rate use and sound statistical thinking depend, in part, on whether problems are or can be represented in [natural frequencies] …."[118] This is a very important concept because it contradicts the findings of several cognitive psychologists who have studied decision making under uncertainty. However, a weakness of this literature is that the majority of research on natural frequencies was performed by Gigerenzer. While he had several collaborators who helped him write various papers on the topic, most of the research built off of his previous work. Only a few other researchers have found similar results.

---

[116] Ibid, 396.

[117] Laura Martignon and Christoph Wassner, "Teaching Decision Making and Statistical Thinking with Natural Frequencies" (paper presented at the International Conference on the Teaching of Statistics 6, South Africa, 2002).

[118] Jonathan J. Koehler, "The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges," *Behavioral and Brain Sciences* 19 (1996): 8.

Another weakness of the literature is that none of the studies used participants that were not physicians or students. Bayesian reasoning has many applications outside of the academic and medical fields and research should incorporate these areas.

While Sedlmeier and Gigerenzer's and Martignon and Wassner's research uncovered excellent results, their research did have one additional weakness. Since they utilized a computerized training program instead of developing a method that would allow people to apply what they learned anywhere, their participants could only use their learned Bayesian reasoning technique if they had the computerized program. While some participants may have developed a general understanding of the concepts and would be able to calculate the math without the program, neither Sedlmeier and Gigerenzer or Martignon and Wassner tested whether or not their participants could perform Bayesian reasoning using their techniques without the computer program.

The weaknesses described above thus leads to the question, is there a non-computerized way to teach Bayesian reasoning to people in disciplines outside of the medical and academic fields? The next section of this chapter will consider this question.

Purpose for the Current Study on Bayesian Reasoning

As discussed earlier in the literature review, Bayesian reasoning is a technique that the intelligence community has been considering for quite some time although it seems to have not yet made its way securely into an analyst's toolbox of techniques. Zlotnick and Schweitzer identified some weakness in the technique, but also provided several benefits of using it. Additionally, the 9/11 Commission Report, the WMD Report, and the Intelligence Reform and Terrorism Prevention Act of 2004 have all called for "improvement in intelligence analysis and the tools and methods that analysts use to create their estimates."[119] As mentioned earlier, the main reason that the intelligence community is shying away from using Bayes is its mathematical complexity. This, however, can be solved using natural frequencies, as Gigerenzer and Hoffrage, Cosmides and Tooby, Sedlmeier and Gigerenzer, and the rest of the researchers on natural frequencies have pointed out.

This thesis combines these two ideas and introduces a method intelligence analysts can use for Bayesian reasoning that incorporates the use of natural frequencies. In order to determine if this new method will be applicable to the intelligence community, this thesis will test the following hypotheses:

Hypothesis One:

> Intelligence-related Bayesian problems worded using natural frequencies will elicit higher Bayesian reasoning amongst intelligence analysts compared to intelligence-related Bayesian problems that are worded using traditional statistical language.

Hypothesis Two:

> A paper and pencil frequency tree method that utilizes natural frequencies can easily be taught to intelligence analysts within 1 ½ hours and elicit a higher level of Bayesian reasoning than Bayesian problems that are only worded using natural frequencies.

[119] Wheaton, Lee, and Deshmukh, "Bayesian Statistics In The Real World."

This study will vary from the previous research in that it offers a new teaching method for Bayesian reasoning that has not already been tested and applies it to a field where limited research on this topic exists. Additionally, research on this topic will contribute to multiple areas of study already discussed in this literature review. The results found in this study will add to the existing research on natural frequencies and human decision making, the development of teaching aids that utilize natural frequencies, and the applicability of Bayesian reasoning in the intelligence field.

CHAPTER 3:

METHODOLOGY

In order to answer the stated hypotheses in the previous chapter, I had to perform three different experiments. The first experiment tested intelligence analysts' performance on Bayesian problems that were worded using natural frequency compared to Bayesian problems worded using traditional statistical language. This experiment was designed to determine if intelligence analysts were able to correctly answer more Bayesian problems when the questions were presented using natural frequencies compared to when they were not and to determine if my first hypothesis was correct. The second experiment built off of the idea of natural frequencies and tested a Bayesian reasoning method using a natural frequency tree format. This second experiment tested the second hypothesis by comparing performance results obtained from using the natural frequency tree method to performance results obtained from wording Bayesian problems using natural frequencies. This experiment was used to determine if using the natural frequency tree method led to better performance on Bayesian problems compared to when problems were just worded using natural frequencies. A third experiment was performed to address some concerns raised about the methodology and results from the second experiment. This experiment was essentially the same as the second experiment, but there were minor changes made to the methodology and instruments used that attempted to correct for the problems. This methodology section will provide details regarding the three different research designs.

Natural Frequency Verification Experiment

Research Design

Before the I could test the validity of and practical use of a Bayesian reasoning method which intelligence analysts could easily use and apply, I had to first determine if intelligence analysts were able to answer Bayesian type questions more successfully when the questions were worded using natural frequencies compared to traditional statistical language. In order to do this, my fellow collaborators[120] and I replicated Gerd Gigerenzer and Ulrich Hoffrage's study on the ability of physicians to perform Bayesian reasoning using both traditional conditional probabilities and natural frequencies.[121] (Details regarding the specific study are located in the previous Literature Review chapter.) This research design utilized a simple post-test experimental design. This type of design is the simplest of all experimental designs.[122] It uses two groups, a control group and an experimental group, both of whose subjects were randomly assigned to control for extraneous variables. Extraneous variables are "other factors besides [the] independent variable that may cause change in the dependent variable."[123] The experimental group is exposed to the independent variable, or stimulus, and the control group is not. The dependent variable, however, is measured for both groups after the testing is completed.

---

[120] Professors Kristan J. Wheaton and Hemangini Deshmukh at Mercyhurst College co-authored the paper "Bayesian Statistics In The Real World Of Intelligence Analysis: Lessons Learned" that was presented at the International Studies Association's 2007 Annual Convention in Chicago, Illinois. This paper was based on this experiment.

[121] Hoffrage and Gigerenzer, "Using Natural Frequencies," 538.

[122] Janet Buttolph Johnson and Richard A. Joslyn, *Political Science Research Methods* (Washington, DC: CQ Press, 1986), 93.

[123] Johnson and Joslyn, *Political Science Research Methods*, 116.

Subjects

To be able to apply the results of the experiment to the intelligence community, we chose to use undergraduate students at the Mercyhurst College Institute For Intelligence Studies studying to become intelligence analysts in place of the physicians Hoffrage and Gigerenzer used for their study. Since many analysts already in the intelligence community require security clearances and are in classified positions, it is very difficult to engage them as a group and we felt that these students best characterized intelligence analysts already in the community.

Mercyhurst College offers the only applied four-year degree program in Intelligence Studies in the world. From the beginning of their first year, the students participate in a rigorous curriculum that prepares them to be analysts in the national security, law enforcement, and business intelligence communities. In addition to the rigorous curriculum, all students within the program participate in real world intelligence projects through internships or simulations of intelligence projects through class assignments, which require them to produce high-quality open-source intelligence products. By the time the students are seniors, they are essentially experienced intelligence analysts without clearances. Therefore, my fellow team members and I relied on ecological validity, which is whether or not the effect on the sample population can be applied to the larger population that it represents.[124]

To control for random assignment, we used two Strategic Intelligence classes of comparable sizes. The Strategic Intelligence class is a capstone class taken by all seniors within the Intelligence Studies program during the fall term. Since all of the students in

---

[124] Reis, Harry T. and Charles M. Judd, eds, *Handbook Of Research Methods In Social And Personality Psychology* (Cambridge: Cambridge University Press, 2000), 12.

the class were seniors, they all had similar experience and ability levels. Combined, 67 senior intelligence students participated in the study.

Process

As part of an ordinary classroom exercise, the professor teaching the two Strategic Intelligence classes asked the students to answer two different questions, both of which presented a hypothetical intelligence situation (see Appendices B and C). As in Hoffrage and Gigerenzer's study, one of the questions incorporated the use of standard statistical terminology and the other used natural frequencies. Both of these questions were structured identically to Hoffrage and Gigerenzer's questions. For each class, the professor gave half of the class the question utilizing statistics and the other half of the class the question that used natural frequencies. The independent variable in this experiment was the use of natural frequencies in the wording of the problem instead of traditional statistical language. The dependent variable in the experiment was the number of people who were able to answer the question correctly, both in the natural frequencies group and the traditional statistical language group.

Data Analysis Procedures

Since the intent of the experiment was to test the reasoning skills of the students and not their mathematics skills, my fellow team members and I scored the answers to the questions using a ten-point margin. We felt that a ten-point margin would account for any math errors but would not compromise the results' indication of a reasoning style.

We then used a Microsoft Excel spreadsheet to record the students' answers to the two questions and to determine the number and percentage of students that answered the questions correctly within the ten-point margin. Once this data was calculated, my fellow team members and I used statistical analysis to determine if the results were significant.

Statistical analysis was performed on two separate sets of data. To determine if there was a statistically significant difference between using a traditional statistics language approach and one utilizing natural frequencies, we used a Z-distribution two-hypothesis test. Since we replicated Hoffrage and Gigerenzer's study, we also compared the two different subject groups (doctors and intelligence analysts) that answered the problems articulated in traditional statistical terms to determine if there were any pre-existing differences between the two groups.

Bayesian Reasoning Using a Natural Frequency Tree Method Experiment

Research Design

As mentioned earlier, the second experiment I performed built off of the previous natural frequency verification experiment. The natural frequency tree method experiment attempted to prove that by using a structured paper and pencil method, intelligence analysts could easily learn how to perform Bayesian reasoning and would be able to correctly answer Bayesian problems using this method better than if the problem were just worded using natural frequencies.

The research design for this experiment also utilized a simple post-test experimental design; however, there was a little modification. While there was both a control and experimental group, I did not actually test a control group. Instead, I used the results from the natural frequency group in the previous study as the control group since this group did not use the frequency tree method to perform Bayesian reasoning.

Since this experiment, unlike the previous experiment, was not part of a classroom exercise, the Mercyhurst College Institutional Review Board (IRB) had to grant me permission before I could conduct the experiment. Any researcher performing research at Mercyhurst College that involves the participation of human beings is required to go through the IRB process. The IRB process ensures that the researcher is not doing anything that will have harmful effects on the participants' health, safety, and emotional or social well being (see Appendix D).

Subjects

All of the participants were students from the Mercyhurst College Institute for Intelligence Studies (MCIIS). As mentioned earlier, from the beginning of their first year, the students participate in a rigorous curriculum that prepares them to be analysts in the national security, law enforcement, and business intelligence communities. Therefore, it makes the most sense to use the students within this program since it is extremely difficult to engage as a group analysts already working in the intelligence field. In addition, these analysts are typically not encouraged to participate in experiments.

I contacted all of the students, both undergraduate and graduate, within MCIIS via email (see Appendix E). The email contained information regarding what the study entailed and the benefits of understanding Bayesian reasoning. In addition to email, I also stopped by all of the classes to personally describe the experiment and answer students' questions. By contacting all of the students within the program, I attempted to obtain a randomly selected group of participants. However, as an incentive to participate, most of the professors within the department offered extra credit for participation. The amount of extra credit varied by professor, as I left it up to the individual professors to determine how much extra credit to offer. Testing took place at the end of the winter term, two weeks before finals. It was the optimal time to use extra credit as an incentive. Offering extra credit may have influenced some students to participate who may not have participated if extra credit had not been offered, thus affecting the extent to which the participants were randomly selected.

In total, 94 students participated in the experiment, with the majority of them being freshman and sophomores.[125] While the majority did come from those two classes, I also obtained participants that were juniors, seniors, and first-year graduate students. Since the participants covered five different educational levels, there was a variety of experience and ability levels amongst them. Below is a breakdown of the participants by year.

[125] I obtained class information through information provided by the students on the consent form. While 94 participants turned in consent forms at the beginning of the experiment, only 92 actually completed the study. Due to the anonymous nature of the experiment, I was unable to determine which students did not complete the study and thus was unable to remove them from the data set regarding the participants' year of study.

Table 3.1: Natural Frequency Tree Method Experiment Participant Breakdown By Class Year	
Class Year	**Number of Participants**
Freshman	47
Sophomore	21
Junior	10
Senior	4
First Year Graduate Students	12
Total Participants	**94**

Process

Before any part of the experiment could commence, I had to obtain permission from all of the participants. When the participants arrived, I had each one sign a participation consent form (see Appendix F). The form gave a general overview of what the participants could expect during the experiment, the time commitment, the guarantee of anonymity, and the assurance that there were no foreseeable risks associated with participation in the study. The consent form also allowed me to obtain some very general demographic information, such as the class year of the participants.

To teach the frequency tree method to the participants, I used a Microsoft PowerPoint presentation. The PowerPoint presentation gave a quick history and overview of Bayesian Reasoning as well as outlined the steps of the frequency tree method and walked through an example problem. Before the presentation began, I provided all of the participants with their own copy of the frequency tree method procedures, which included an example problem (see Appendix G). The procedures and example problem were identical to the information outlined in the PowerPoint presentation. During the presentation, I emphasized the points that I felt would cause the most confusion and tried to clarify them as much as possible. I also allowed for any questions that arose.

After going through the entire PowerPoint presentation, I handed out a practice question (see Appendix H). The key aspect of the frequency tree method is that it allows for movement around the frequency tree and does not require that the same type of information be present for the method to work. The practice question was set-up similar to the example problem given with the frequency method procedures; however, the type of data given was different. This required the participants to see if they were able to apply the method's concepts. Once the participants were finished with the practice question, I showed them how to correctly apply the frequency tree method and answered their questions.

Once all of the participants' questions were answered, I handed out four test questions (see Appendix I). The purpose of the test questions was to see how well the participants were able to apply what they had learned about the frequency tree method. The first problem was set-up exactly like the example problem. In the test questions following the first question, the type of information given was varied and the problems became harder. The last question had additional information and required the participants to identify the relevant data. For all of the test questions, the participants were allowed to refer to the frequency tree method procedures handed out at the beginning of the experiment.

Following the completion of the test questions, the participants were asked to fill-out a short questionnaire for qualitative feedback (see Appendix J). The questions asked the participants to rate their understanding of Bayesian reasoning prior to the experiment and afterwards, how important they felt a mathematics background was in order to be able to understand and perform the frequency tree method, and how useful they felt the frequency tree method would be for an analyst. Once a participant turned in his or her questionnaire, I gave him or her a debriefing statement that explained the purpose of the study and how the results would be used (see Appendix K).

In this experiment, the independent variable was the frequency tree method that the experimental group used. The dependent variable for both the control group and the experimental group was the number of participants who correctly answered the Bayesian problem.

Data Analysis Procedures

Just as in the previous experiment, the intent of the frequency tree method experiment was to test the reasoning skills of the students and not their mathematics skills; therefore, I once again scored the answers to the questions using a ten-point margin. Using a ten-point margin for these results was also required in order for me to compare the results from the experimental group to the results from the control group.

Once again, I used a Microsoft Excel spreadsheet to record the participants' answers to the four questions and to determine the number and percentage of participants that answered the questions correctly within the ten-point margin. In addition to recording the results from the test questions, I also recorded the participants' answers to the questionnaire. By using and Excel spreadsheet, I was able to determine the average answers for all of the questionnaire questions. Once, all of this data was calculated, I used statistical analysis to determine if the results were significant.

Statistical analysis was performed on two separate sets of data. To determine if there was a statistically significant difference between using the frequency tree method to solve Bayesian problems and Bayesian problems worded using natural frequencies, I used a one-sample Z-distribution hypothesis test. I also compared answers from the questionnaire to determine if there was any significance between the participants' average understanding of Bayesian reasoning prior to using the frequency tree method and their average understanding of Bayesian reason after learning how to use the frequency tree method.

Revised Natural Frequency Tree Method Experiment

Research Design

As mentioned earlier, the third experiment I performed was essentially the same as the previous natural frequency tree experiment; however there were minor changes made the methodology and instruments utilized. The results from the previous experiment (which will be discussed in the next chapter) identified a mistake made by more than half of the participants. As part of the natural frequency tree method, participants are required to determine the reference population for the very top node (see figure on next page). The size of reference population depends on the information provided in the problem. Thus, the reference population's size varies by problem and is never just one specific number. However, more than half of the participants entered 100 as the reference population when the reference population should have been a different number, thus leading to incorrect answers.

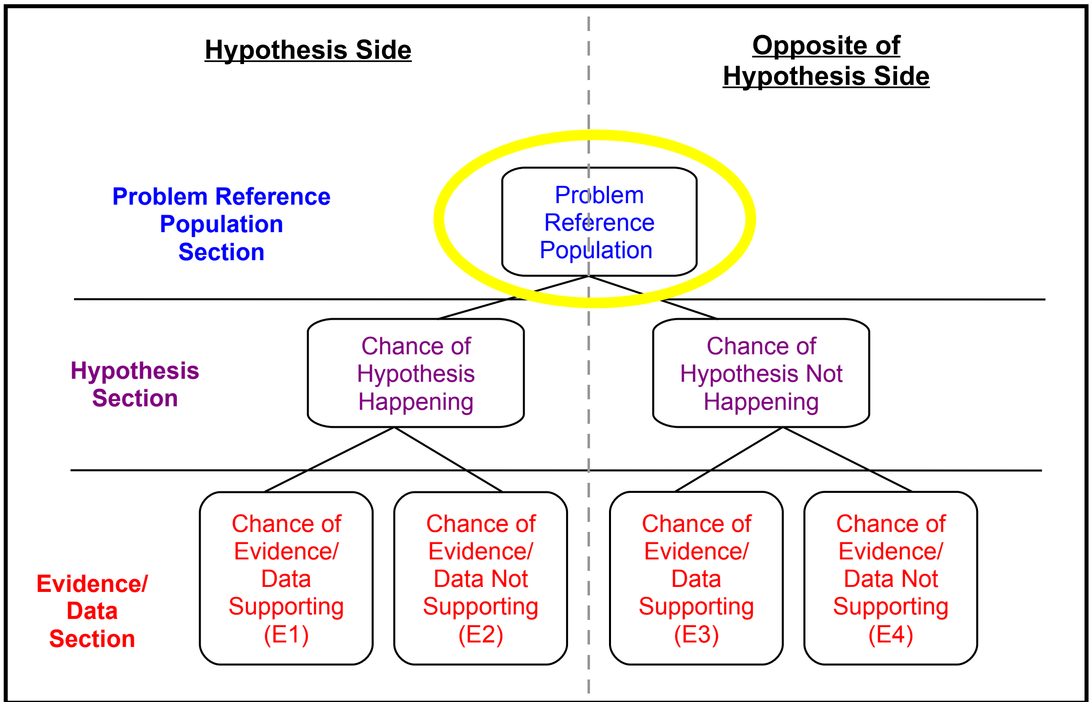


Figure 3.1: Natural Frequency Tree Representation With Reference Population Node Identified.

From reviewing the methodology I used in the previous experiment, I hypothesized that the reason a majority of the participants put 100 in the reference population node when they were not supposed to was because they used the practice question (see Appendix H), that had a reference population of 100, as their example to follow when solving the test questions. Even though throughout the experiment I emphasized that the reference population can be any number, I felt that because of the participants had an example problem right in front of them to follow when solving the test questions, they did not utilize the frequency tree method instructions (see Appendix G) that explained how to determine the reference population based on the information given in the problem. The revised natural frequency tree method experiment attempted to prove that this was the reason for the participants' poor performance using the natural frequency tree method.

The research design for this experiment also utilized a simple post-test experimental design and was similar to that of the previous natural frequency tree method experiment. Instead of using the natural frequency group from the natural frequency verification experiment (the first experiment) as my control group, I used the group from my previous natural frequency tree method experiment since they were exposed to the original methodology and not to any of the changes made for this follow-up experiment.

Since this experiment, was similar to the previous natural frequency tree method experiment, and because there were very few changes made to the original methodology, I did not need to obtain a second approval from the IRB to perform this second natural frequency tree method experiment.

Subjects

Once again, all of the participants were students from within the MCIIS program. However, my available pool of participants this time was limited to those who had not participated in my previous experiment. I needed to use students who had not participated in my previous experiment because it was necessary for me to be able to compare the participants' results from this study to the participants' results in the previous study. For this to be possible, I had to ensure my participants were not already familiar with my natural frequency tree method.

I contacted all of the students, both undergraduate and graduate, within MCIIS again via email (see Appendix L). This email was similar to my previous email in that it contained information regarding what the study entailed and the benefits of understanding Bayesian reasoning. Additionally, this email emphasized my need for students who had participated in my previous study. By contacting all of the students within the program, I attempted to obtain a randomly selected group of participants from those who had not previously participated. However, as an incentive to participate, most of the professors within the department once again offered extra credit for participation. As before, the amount of extra credit varied by professor, as I left it up to the individual professors to determine how much extra credit to offer. Since offering extra credit may have influenced some students to participate who may not have participated if extra credit had not been offered, the extent to which the participants were randomly selected may have been affected.

In total, 11 students participated in the experiment, with the majority of them sophomores.[126] Below is a breakdown of the participants by year who participated in this second experiment.

[126] Once again, I obtained class information through information provided by the students on the consent form.

Table 3.2: Revised Natural Frequency Tree Method Experiment Participant Breakdown By Class Year	
Class Year	**Number of Participants**
Freshman	0
Sophomore	9
Junior	2
Senior	0
First Year Graduate Students	0
Total Participants	**11**

Process

The process I used for this second experiment was almost identical to my previous natural frequency tree method experiment (see previous section for details). There were a few minor changes in the Microsoft PowerPoint presentation as well as the frequency tree method procedures. Most of the changes involved wording; however, there was one change made to one of the illustrations (see Appendix M for changes).

Since I felt that the practice question I used in the previous experiment was the cause for the confusion and misuse of the number 100 in the reference population, I changed the information given in the practice question in order to eliminate the use of 100 as the reference population (see Appendix N). This way, the participants would not have an example that showed 100 as the reference population. I also received some feedback from my previous experiment stating that more practice questions would have been helpful. I decided to add a second practice question that did not require 100 as the reference population (see Appendix O).

Instead of having the participants try the practice question on their own, I walked through the first practice question with them. I asked them to derive from the practice question the information necessary to fill-in the natural frequency tree. Finding the solution to this practice question was a group effort. For the second practice question, I asked them to try and fill-in the natural frequency tree on their own. Once the participants were finished with this second practice question, I showed them how to correctly apply the frequency tree method. I hoped that this new approach to the practice problems would eliminate the confusion that I noticed the participants in the first experiment experienced. The rest of the experiment followed the same procedures as the previous natural frequency tree method experiment.

In this experiment, the independent variable was the revised frequency tree method that the experimental group used. Once again, the dependent variable for both the control group and the experimental group was the number of participants who correctly answered the Bayesian problem.

Data Analysis Procedures

I collected and recorded the results for this experiment the same way that I did for the previous natural frequency tree method experiment. As in the previous two experiments, I scored the answers to the questions using a ten-point margin as it was required for me to be able to compare the results from the experimental group to the results from the control group.

Once again, I used a Microsoft Excel spreadsheet to record the participants' answers to the four questions and to determine the number and percentage of participants that answered the questions correctly within the ten-point margin. I also recorded the participants' answers to the questionnaire in this same Excel spreadsheet. As before, I was able to determine the average answers for all of the questionnaire questions. Once, all of this data was calculated, I used statistical analysis to determine if the results were significant.

Statistical analysis was performed on two separate sets of data. To determine if there was a statistically significant difference between using the revised frequency tree method to solve Bayesian problems and the first version of my frequency tree method, I used an independent two-sample t-distribution hypothesis test. I also compared the results from using my revised frequency tree method to the results from my first experiment that tested how well participants were able to Bayesian problems that were just worded using natural frequencies. Finally, I compared the answers from the questionnaire to determine if there was any significance between theses participants' average understanding of Bayesian reasoning prior to using the frequency tree method and their average understanding of Bayesian reason after learning how to use the frequency tree method. I wanted to see if I was able to reproduce the results I found in my previous natural frequency tree method experiment.

The results from the three experiments described above will be discussed in detail in the following results chapter.

CHAPTER 4:

RESULTS

The results from the three natural frequency experiments provide interesting findings regarding the impact natural frequencies have on people's abilities to solve intelligence-related Bayesian problems. This chapter will discuss in detail the findings from the three experiments described in the previous chapter. For all three experiments, statistical analysis was performed to determine if the findings uncovered were significant. For all statistical analysis, Statistical Package for the Social Sciences (SPSS) was used.

Natural Frequency Verification Experiment Findings

As discussed in the previous chapter, the natural frequency verification experiment sought to test the hypothesis that intelligence analysts were able to correctly answer more Bayesian problems when the questions were presented using natural frequencies compared to when they were worded using traditional statistical language. The experiment was a replication of Ulrich Hoffrage and Gerd Gigerenzer's doctor experiment.[127] In the natural frequency group, 26 of the 33 (79 percent) students fell within ten points of the correct answer,[128] while in the traditional statistics group, only 6 of the 34 (18 percent) students were within ten points of the correct answer (see Figures 4.1 and 4.2 on next page).

---

[127] See the Literature Review chapter for a detailed description of Hoffrage and Gigerenzer's experiment.
[128] As mentioned in the Methodology chapter, the intent of the experiment was to test reasoning skills, not mathematic skills. We felt that a ten-point margin would account for any math errors but would not compromise the results' indication of a reasoning style.

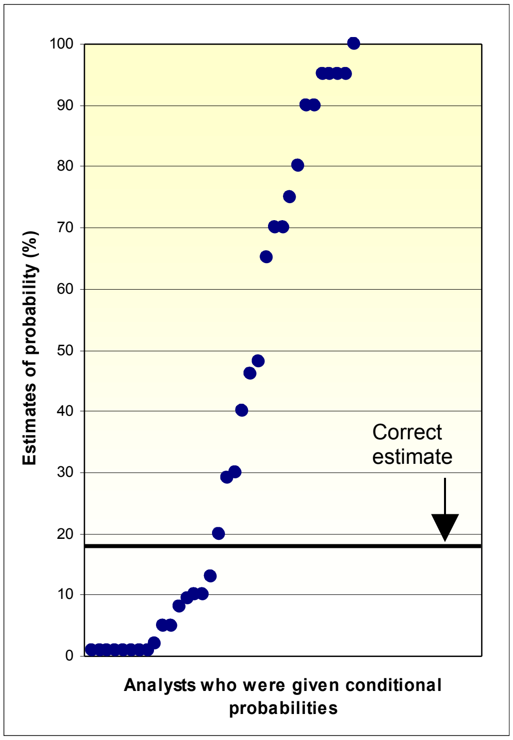


Figure 4.1: Traditional Statistics Group Answers

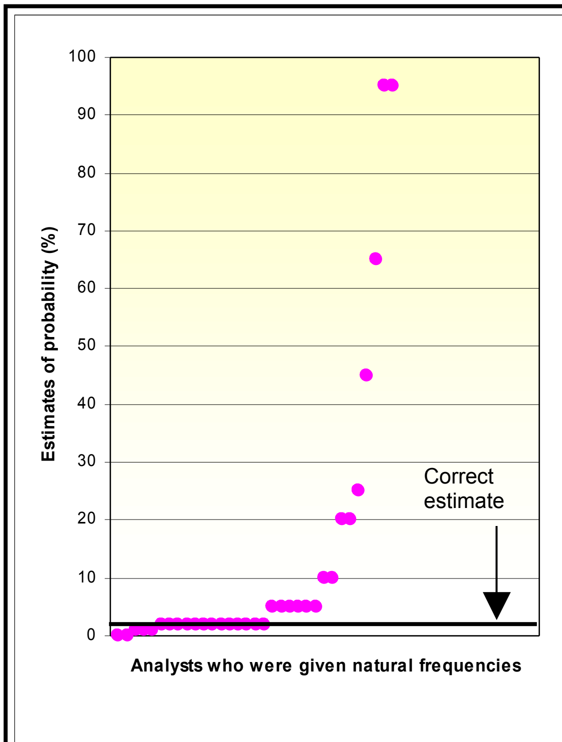


Figure 4.2: Natural Frequency Group Answers

Using a Z-distribution two-hypothesis test[129], it was determined that there was a statistically significant difference between the traditional statistics group and the natural frequency group at a significance level of 5 percent ($\alpha = 0.05$), indicating that these results would occur by chance less than 5 percent of the time. The Z-distribution two-hypothesis test compares a claim between two large sample populations ($n \geq 30$) that have a normal distribution to determine if the difference between them is significant.

For there to be a statistically significant difference between the two groups, the P-value (probability value) must be smaller than the level of significance. In this case, the P-

[129] SPSS does not have a Z-test, but if the degrees of freedom are larger than 30, a t-distribution will approach a Z-distribution. Thus, it is OK to use a t-test in SPPS for a Z-distribution test on large sample populations.

value was equal to 0.002, which was less than the selected significance level of 0.05 (see Table 4.1 below.)[130]

Table 4.1: Natural Frequency Verification Experiment Independent Samples Test Results For The Z-Distribution Two-Hypothesis Test							95% Confidence Interval of the Difference	
		t (Z-test value)	Degrees of Freedom	Sig. (2-tailed) (P-value)	Mean Difference	Std. Error Difference	Lower	Upper
Frequency per Group	**Equal variances not assumed**	**3.183**	57.548	**0.002**	24.87763	7.81542	9.23076	40.5245

The results from this test strongly suggest that using natural frequencies leads to intelligence analysts being better Bayesians, thus supporting the hypothesis that wording problems using natural frequencies leads to better performance amongst intelligence analysts when compared to problems that are worded using traditional statistical language. Additionally, when the natural frequency group's results were compared to Hoffrage and Gigerenzer's findings, they matched exactly. Hoffrage and Gigerenzer found that 19 out of the 24 doctors (79 percent) fell within 10 points of the correct answer (see Figure 4.3 on the next page).[131]

Since these results replicated Hoffrage and Gigerenzer's results so well, another test was used to make sure that there were not any pre-existing differences between the two subject groups (doctors and analysts). To do this, a t-distribution independent two-

[130] For a complete set of the results as well as the raw data for this experiment, see Appendix P.
[131] Image used without permission. Taken from Gerd Gigerenzer and Adrian Edwards, "Simple tools for understanding risks: from innumeracy to insight," *BMJ* 327 (2003): 742, http://www.bmj.com/cgi/reprint/327/7417/741.

hypothesis test was used to compare Hoffrage and Gigerenzer's traditional statistics group to the analyst traditional statistics group. A t-distribution test had to be used instead of a Z-distribution test because Hoffrage and Gigerenzer's sample population was small (less than 30).[132]

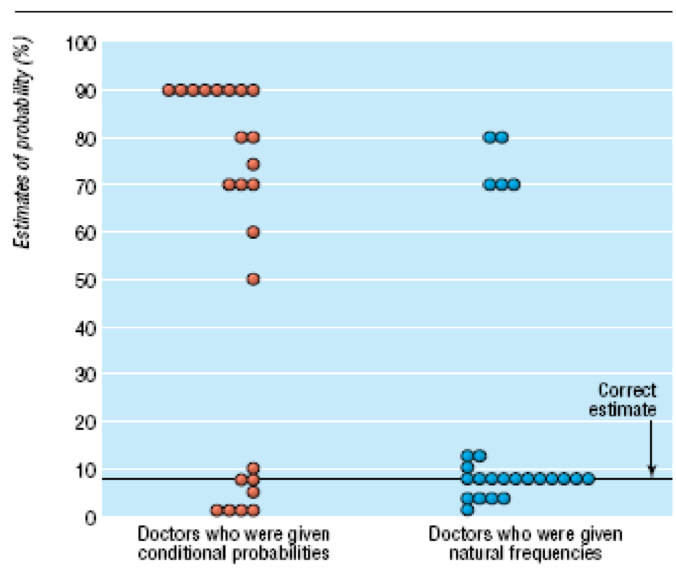


Fig 1 Doctors' estimates of the probability of breast cancer in women with a positive result on mammography, according to whether the doctors were given the statistical information as conditional probabilities or natural frequencies (each point represents one doctor)[2]

Figure 4.3: Results From Hoffrage And Gigerenzer's Natural Frequency Doctor Experiment

While there was a difference between the two traditional statistics groups regarding the number of subjects who came within 10 points of the correct answer (33 percent of doctors vs. 18 percent of analysts), the test results showed that there was no significant difference between the two groups at the 5 percent significance level ($\alpha = 0.05$). The P-

[132] See Appendix P for statistical data regarding the normality assumption check for Hoffrage and Gigerenzer's traditional statistics doctor group. This data also shows the results form Levene's test for the equality of variances. The normality assumption check was required before we could use the t-distribution independent two-hypothesis test since this test can only be used to compare two normal samples. Levene's test was required in order to determine which t-test value should be used.

value this time was 0.108, which was larger than the selected significance level of 0.05 (see Table 4.2 below).[133] These results indicate that analysts graduating from the Mercyhurst College Institute for Intelligence Studies (MCIIS) program are at least as comfortable as doctors with statistically based reasoning.

Table 4.2: Doctors vs. Analysts Independent Samples Test Results For The t-Distribution Independent Two-Hypothesis Test

							95% Confidence Interval of the Difference	
		t (t-test value)	Degrees of Freedom	Sig. (2-tailed) (P-value)	Mean Difference	Std. Error Difference	Lower	Upper
Frequency per Group	Equal variances assumed	**1.634**	56	**0.108**	16.41176	10.04361	-3.7080	36.53153

Bayesian Reasoning Using a Natural Frequency Tree Method Experiment Findings

As indicated in the previous chapter, the purpose of the natural frequency tree method experiment was to test the hypothesis that a paper and pencil natural frequency tree method would lead to better performance by intelligence analysts on Bayesian problems when compared to intelligence analysts' performance on Bayesian problems that were worded using natural frequencies. To determine if this hypothesis was true, the participants' performance results from the first test question in the natural frequency tree verification experiment were compared to the results of the natural frequency group from the previous natural frequency verification experiment using a one-sample Z-distribution hypothesis test. A Z-distribution test was utilized because both of the sample populations were considered to be large ($n \geq 30$).

[133] For a complete set of the results, see Appendix P.

The raw data findings showed that 70 out of the 92 participants (76 percent) correctly answered test question number one within 10 points.[134] When compared to the results from natural frequency group in the previous experiment, which indicated participants were able to calculate the correct answer 79 percent of the time within 10 points, it was determined that the results were not statistically significant at a 5 percent significance level (α = 0.05). The P-value (0.516) from this test was greater than the selected significance level of 0.05(see Table 4.3 below).

Table 4.3: Original Natural Frequency Tree M One-Sample Test Results For The Z-Distribu Hypothesis Test

	Test Value = 0.79					
					95% Confidence Interval of the Difference	
	t (Z-test value)	Degrees of Freedom	Sig. (2-tailed) (P-value)	Mean Difference	Lower	Upper
Test Question Number 1	**-0.651**	91	**0.516**	-0.02913	-3.70800	36.53153

[134] To view the complete set of raw data for this experiment and more detailed statistical results, see Appendix Q.

Since the results were not significant, the findings indicate that the paper and pencil natural frequency tree method does not encourage better performance than if analysts were to use their own form of calculation to solve an intelligence-related Bayesian problem that was worded using natural frequencies, which does not support the initial hypothesis that the paper and pencil natural frequency tree method would yield better results than wording problems using natural frequencies.

While analyzing these results, several hypotheses came to mind as to why the performance using the paper and pencil natural frequency tree method was lower than the performance on Bayesian problems worded using natural frequencies. One hypothesis was that the lack of formal statistical training amongst the freshman and sophomore classes led to a lower performance. The rationale for this hypothesis was that since the majority of the sample population was made up by these two classes, there was the chance that their lack of experience had a significantly negative impact on the experiment's results. According to the post-test questionnaire,[135] the students felt that on average a mathematical background would be useful for performing the frequency tree method (the average answer was 2.84 on a scale from 1 – 5, with 1 being not useful at all and 5 being extremely useful). This hypothesis seemed like a fairly reasonable possibility.

However, research into the student's mathematical backgrounds proved this hypothesis wrong. Out of the 66 combined freshman and sophomore participants, only

[135] See Appendix R for complete post-test questionnaire results.

29 had not yet taken a statistics class (see Table 4.4 below).[136] Additionally, according to the post-test questionnaire, the students indicated that on average, they had a higher than average math background (the average answer was 3.25 on a scale from 1 – 5). It was thus assumed that this was not the reason for less than expected performance using the natural frequency tree method since a majority of the participants had taken statistics and they considered themselves to have a fairly strong mathematical background.

Table 4.4: Status of Statistics Education Amongst Freshman and Sophomore Participants	
Status of Statistics Education[137]	**Number of Students**
Have not yet taken a statistics class.	29
Took a statistics class prior to the fall term.	9
Took a statistics class during the fall term.	10
Took a statistics class during the winter term.	18
Total Freshman and Sophomore Participants	**66**

A second hypothesis for the lower than expected results was that the maturity level of the participants negatively impacted the experiment's results. As mentioned earlier, the majority of participants for the natural frequency tree experiment were freshmen and sophomores. Freshmen and sophomores have a different maturity level than seniors, who made up the sample population of the natural frequency group in the natural frequency verification experiment. By the time a student is a senior, he or she has decided that intelligence is the field he or she wants to work in and is likely to consider

[136] An Introduction to Statistics class is required for all students in the MCIIS program.
[137] As mentioned in the Methodology chapter, the experiment was given during the winter term. The status of the students is based on their education level at the time of the experiment.

learning an analytical technique more seriously. Additionally, seniors will have had an internship experience in a professional environment, which also leads to a higher maturity level. Freshmen and sophomores are still at the point in their education where they are experimenting with different areas of study to determine what field interests them the most and may not take learning as seriously. Unfortunately, further study into this hypothesis could not be done.

Another hypothesis that came to mind was the difference between the settings of the natural frequency verification experiment and the natural frequency tree experiment. The natural frequency verification experiment was done as part of a class and was led by a professor compared to the natural frequency tree experiment that took place outside of class and was led by a graduate student. The higher performance elicited by the natural frequency group in the natural frequency verification experiment may have resulted due to students taking it more seriously because it was part of a classroom exercise and not considered an "experiment." Also, the students who participated in the "experiment" outside of class were also offered extra credit. Since extra credit was not awarded based on the student's performance, the participants may not have taken the experiment outside of the classroom as seriously. Additionally, the presence of a professor during the natural frequency verification experiment may have made a significant difference. Unfortunately, this hypothesis was not able to be tested.

A final hypothesis that attempted to explain the statistically equivalent results using the natural frequency tree method when compared to Bayesian problems worded using natural frequencies was the misuse of 100 as the reference population for the entire natural frequency tree. Out of the 92 participants, 58 (63 percent) put 100 in the reference population node (very top node of the frequency tree) incorrectly at least once. Out of the four test questions, only one question's reference population should have been 100. In order to determine whether or not this could be the cause, the revised natural frequency tree method experiment was performed. The findings from that experiment are described in the next section.

Some positive findings can be found in the post-test questionnaire results. The first two questions on the questionnaire asked the participants to rate their understanding of Bayesian reasoning prior to learning the natural frequency tree method and after learning the natural frequency tree method. On a scale of 1-5, with 1 being extremely low and 5 being extremely high, the participants rated their prior knowledge on average as a 1.19 and their knowledge after the experiment on as a 3.15. To determine if the difference between the two answers was significant, a t-test for dependent or paired samples was performed.[138] This test was utilized because both of the samples were considered to be normally distributed since their sample sizes were large ($n \geq 30$). In this case, the sample population was 93. The t-test for the paired samples showed that there was a significant difference between the participants' knowledge level of Bayesian reasoning prior to learning the paper and pencil frequency tree method and after learning the method. The P-value for this test was 0.00, which was less than the selected 5 percent ($\alpha = 0.05$) significance level (see Table 4.5 on next page.) While the paper and pencil natural frequency tree method did not show a significant increase in participants' performance using Bayesian reasoning, the results indicate that the method did improve the participants' understanding of Bayesian reasoning. In fact, on average, the participants rated this method as being very useful (3.63 on a scale from 1 – 5, with 1 not

[138] See Appendix R for complete statistical results.

useful and 5 being very useful) for an intelligence analyst, and 73 percent of the participants said that they plan to use this method of Bayesian reasoning as an analytic tool thus indicating that the natural frequency tree method increased their confidence in using Bayesian reasoning. This was further supported by some of the comments the participants provided on the post-test questionnaires. One participant commented, “I thought it was very easy to understand, more so than a very complicated formula.” Another participant said, “I thought this approach seemed much easier than the actual Bayesian reasoning equation.” Finally, a third participant said, “This is a much simplified version of Bayesian reasoning and [illegible] time in the analytic field.”

Table 4.5: Original Natural Frequency Tree Method Post-Test Questionnaire Paired Samples Test Results For The t-Test For Paired Samples

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t (t-test value)	Freedom Degrees of	Sig. (2 -tailed) (P-value)
Pair 1	**Questionnaire Question 1 - Questionnaire Question 2**	-1.95	0.915	0.095	-2.14	-1.26	**-20.6**	92	**0.000**

Revised Natural Frequency Tree Method Experiment Findings

The purpose of this experiment was to determine if the changes made to the methodology of the original paper and pencil natural frequency tree method answered the hypothesis regarding the use of 100 in the reference population node by resulting in improved participant performance when compared to the results from the first natural frequency tree method and when compared to the results of the natural frequency verification experiment. In order to determine if there was a significant improvement in performance resulting from the changes made to the natural frequency tree methodology, an independent two-sample t-distribution hypothesis test was performed to compare the results from the first test questions in both experiments.[139]

A t-distribution test was used instead of a Z-distribution test due to the small sample size ($n = 11$) from the revised natural frequency tree experiment.[140] When the findings from the revised natural frequency tree experiment were compared to the findings from the initial natural frequency tree experiment, it was determined that there was not a significant improvement in performance at the 5 percent level ($\alpha = 0.05$), thus failing to support the hypothesis that the changes made to the natural frequency tree method would lead to improved performance when compared to the results from the original natural frequency tree experiment. The P-value returned for this test was 0.3375, which is higher than the selected significance level of 0.05 (see Table 4.6 below).

[139] Before comparing the results from the experiments, a normality check had to be done on the sample population from the revised natural frequency tree experiment because of its small size. The normality assumption was met using Normal P-P plots and Levene's test determined if equal variances were assumed or not assumed. For more details regarding these tests and the complete statistical results from this experiment, see Appendix S.

[140] A sample size is considered small if ($n < 30$).

Table 4.6: Independent Samples Test Results For The t-Distribution Two-Hypothesis Test For The Revised Natural Frequency Tree Method Experiment And Original Natural Frequency Tree Method Experiment							95% Confidence Interval of the Difference	
		t (t-test value)	Degrees of Freedom	Sig. (1-tailed) (P-value)	Mean Difference	Std. Error Difference	Lower	Upper
Frequency per Group	Equal variances assumed	**-0.42**	101	**0.3375**	-0.05731	0.13608	-0.3273	0.21263

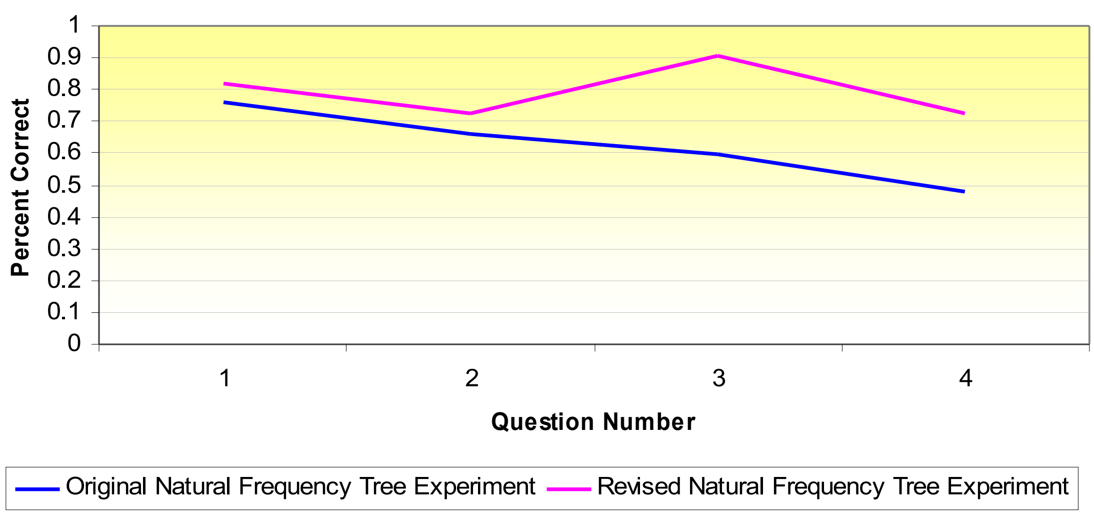
While the raw data shows that the percentage of participants in the revised natural frequency tree experiment who obtained the correct answer for test question one within ten points (approximately 82 percent)[141] is higher than the percentage of participants in the original natural frequency tree experiment (approximately 76 percent), the statistical analysis results indicate that the changes made to the natural frequency tree methodology were within statistically equivalent limits.

In addition to comparing the first test questions from both experiments, the trend lines derived from the four test questions in both experiments were compared as well. The trend line for the first natural frequency tree experiment shows a steady decline in performance over the four different test questions, which is not surprising as the test questions gradually became more difficult; however, the trend line for the four test questions in the revised natural frequency tree experiment shows a different result (see Figure 4.4 below).[142] The trend line from this experiment shows less of a decline in

[141] See Appendix S for the complete findings and raw data for this experiment.
[142] See Appendices Q and S for actual numbers.

performance and in fact shows an increase in performance for test question number three. In order to determine if there was a significant difference between the two trend lines, an independent two-sample t-distribution hypothesis test was performed.

Figure 4.4: Trend Lines For Test Questions 1 Through 4 For The Original Natural Frequency Tree Experiment And The Revised Natural Frequency Tree Experiment



The results from the independent two-sample t-distribution hypothesis test showed that there was not a significant difference between the two different trend lines at the 5 percent significance level ($\alpha = 0.05$), which does not support the hypothesis that the changes made to the natural frequency tree method significantly improved overall performance. However, in this test, the P-value was equal to 0.055, which was barely higher than the selected significance level of 0.05 (see Table 4.7 below).[143] While these results are not significant at a 95 percent confidence interval, they are significant at a 94.5 percent confidence interval, which indicates that there is a 94.5 percent chance that the difference between the two different trend lines was not due to chance. This is a very

[143] See Appendix T for complete statistical findings, including normality tests.

important observation because it indicates that the changes made to the natural frequency tree experiment did improve performance considerably, just not at the 5 percent significance level.

Table 4.7: Trend Lines Independent Samples Test Results For The t-Distribution Independent Two-Hypothesis Test

							95% Confidence Interval of the Difference	
		t (t-test value)	Degrees of Freedom	Sig. (2-tailed) (P-value)	Mean Difference	Std. Error Difference	Lower	Upper
Frequency per Group	Equal variances assumed	**-2.37**	6	**0.055**	-0.17250	0.07267	-0.3503	0.00532

These findings were also compared to Gigerenzer and Sedlmeier's computerized natural frequency tree method immediate transfer findings. The results from both paper and pencil natural frequency tree experiments show that participants were not able to generalize from the training problems to the test questions as well as Gigerenzer and Sedlmeier's subjects. When Gigerenzer and Sedlmeier had their participants apply the knowledge they gained from the training problems to transfer problems (which are equivalent to the test questions in the paper and pencil natural frequency tree experiments), they found that their subjects who learned their computerized natural frequency tree method were able to answer 90 percent of the problems correctly.[144] The results from the two paper and pencil frequency tree method experiments done in this

[144] Sedlmeier and Gigerenzer, "Teaching Bayesian Reasoning," 399.

thesis show that students were only able to provide at best 80 percent correct Bayesian answers, which were the results from the revised natural frequency test experiment.[145] The transfer ability of the participants from the original natural frequency tree method, which had a significantly larger sample population, was only 63 percent. These results show that the paper and pencil natural frequency tree method for Bayesian reasoning introduced in this thesis do not offer as much knowledge transfer as computerized natural frequency tree methods. While both of the methods, Sedlmeier and Gigerenzer's computerized training program and the paper and pencil method introduced in this thesis, were taught in less than two hours, the computerized training program was able to provide more individual feedback to participants, which is the likely reason why the transfer rates were considerably higher.

Finally, the findings from this second natural frequency tree experiment were compared to the results from the natural frequency verification experiment to determine if this revised paper and pencil natural frequency tree method would lead to better performance by intelligence analysts on Bayesian problems when compared to intelligence analysts' performance on Bayesian problems that were just worded using natural frequencies. In order to determine if there was a significant difference in performance between the two participant groups, a one-sample t-distribution hypothesis test was performed.[146] The results from this test show that once again that there was not a significant difference in performance at a 5 percent significance level ($\alpha = 0.05$) between those using the natural frequency tree method to solve an intelligence-related Bayesian

---

[145] The 80 percent was determined by adding up the correct number of answers to the test questions in the revised natural frequency tree experiment and dividing them by the number of test questions. The results for the original natural frequency tree experiment were lower at 63 percent. See Appendix U for more information.

[146] Once again, a normality check had to be performed in order to use this hypothesis test. See Appendix S for the complete normality check results and the complete statistical results from this experiment.

problem and those who solved an intelligence-related Bayesian problem that was worded using natural frequencies. The P-value from the one-sample t-distribution hypothesis test was 0.822, which was higher than the selected significance level of 0.05 (see Table 4.7 below).

Table 4.8: One-Sample Test Results For Th Distribution Hypothesis Test For The Re Frequency Tree Method Experiment And N Verification Experiment

	Test Value = 0.79					
					95% Confidence Interval of the Difference	
	t (Z-test value)	Degrees of Freedom	Sig. (2-tailed) (P-value)	Mean Difference	Lower	Upper
Test Question Number 1	**0.231**	10	**0.822**	0.02818	-0.2436	0.2999

These results indicate that the revisions made to the natural frequency tree method did not significantly improve the method to a point where performance on Bayesian problems was better using the paper and pencil natural frequency tree method compared to just wording Bayesian problems in natural frequencies, which does not support the hypothesis.

Results, however, from the post-test questionnaire once again show a significant improvement at the 5 percent significance level ($\alpha = 0.05$) of the participants' understanding of Bayesian reasoning after learning the natural frequency tree method. On a scale of 1-5, with 1 being extremely low and 5 being extremely high, the participants rated their prior knowledge of Bayesian reasoning on average as a 1.27 and their knowledge after the experiment on as a 3.27.[147] To determine if the difference between the two answers was significant, a t-test for dependent or paired samples was performed.[148] Since the P-value returned from this test was 0.000 and less than the selected significance level of 0.05, the difference between the two questions is considered to be significant (see Table 4.8 below).

Table 4.9: Paired Samples Test Results For The t-Test For Paired Samples (Revised Natural Frequency Tree Experiment)

	Paired Differences							
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t (t-test value)	Freedom Degrees of	Sig. (2 -tailed) (P-value)

[147] See Appendix V for the complete set of post-test questionnaire data.
[148] See Appendix V for the normality assumption checks and complete statistical findings.

Pair 1	**Questionnaire Question 1 - Questionnaire Question 2**	-2.00	1.095	0.330	-2.74	-1.26	**-6.06**	20	**0.000**

While the overall statistical results from the revised natural frequency tree experiment indicate that there was not an increase in performance that resulted from changes made to the methodology, the incorrect use of 100 as the reference population in the top node did not occur once. This positive finding shows that the changes to the original natural frequency tree method did correct one of the misunderstandings regarding the use of the method. Additionally, anecdotal evidence indicates that participants in the revised natural frequency tree experiment seemed to have a better understanding of the method as they were able to correctly answer more questions during the practice problems than the participants in the first natural frequency tree experiment.

Natural Frequency Tree Method Questionnaire Findings

The post-test questionnaire findings from both natural frequency tree experiments also provided some important qualitative feedback regarding the natural frequency tree method and training sessions. Feedback from the original natural frequency tree experiment indicates that on average, the participants felt that the natural frequency tree method was easy to follow (they rated the method as 3.16 on a scale from 1 – 5, with 1 being not easy and 5 being extremely easy) and that the training session was very useful in explaining the method and answering participants' questions (the participants rated the usefulness of the training session on average as 3.63 on a scale from 1 – 5, with 1 being not useful and 5 being extremely useful).

The findings from the revised natural frequency tree experiment post-test questionnaire are also positive. The participants in this experiment felt that the method was very easy to follow (their average answer was 3.55) and that the training session was also very useful in explaining the method and answering their questions (their average answer was 3.73). The findings from the revised natural frequency tree experiment are higher than those from the original natural frequency tree experiment indicating that the changes made to the methodology and training session improved participants' understanding of how the method worked and also made the training session more useful.

Summary Of Results

The findings discussed in the previous sections indicate that natural frequencies make it easier for intelligence analysts to answer Bayesian problems. This was shown by replicating Hoffrage and Gigerenzer's doctor experiment. The results from this experiment led to the development of the original natural frequency tree method in hopes of discovering a paper and pencil method that would result in better performance when compared to wording problems using natural frequencies. However, statistical analysis showed that there was not a significant difference between participants' performance using the natural frequency tree method and their performance when answering questions worded using natural frequencies. This result led to several hypotheses, one of which was tested to see if performance improvement could be made. The second natural frequency tree experiment attempted to elicit higher performance by making modifications to the original natural frequency tree method. The performance resulting from this experiment was not statistically significantly better; however, the findings did indicate that improvements made to the natural frequency tree method can reduce confusion and possibly lead to better performance.

There are still some more changes that need to be made to the methodology to ensure that those learning the method have a complete understanding of the concepts involved and how the method works. The next chapter will offer some recommendations for additional improvements and changes to the methodology that were not able to be made to the experiments performed for this thesis.

CHAPTER 5:

CONCLUSIONS

The purpose of this study was to determine whether or not the natural frequency tree method introduced in this thesis would lead to better results when compared to Bayesian problems that were worded using natural frequencies. Through the use of multiple experiments and statistical analysis, it was determined that statistically, the paper and pencil natural frequency tree method does not yield as high as performance as wording problems using natural frequencies. However, it can not be concluded that overall the natural frequency tree method was not successful. The two experiments were different in terms of the types of problems used, the methodologies, and the participants. The natural frequency verification experiment's participants were college seniors and the experiment only used one Bayesian problem that was fairly simple. Additionally, the experiment was part of a class and was led by a professor. The paper and pencil natural frequency experiments were more robust as they involved four Bayesian problems of varying degrees and used participants with a variety of education levels. While the overall statistical findings of the natural frequency tree experiments were disappointing, other important and more positive findings were found.

One important conclusion resulting from this study is that there is an interest amongst future intelligence analysts to learn about Bayesian reasoning. The large turnout for the original natural frequency experiment indicates this and shows that the mathematical component of Bayesian reasoning did not turn analysts away. Additionally, as discussed in the previous results chapter, the participants indicated that they felt the natural frequency tree method to be useful as an analytic tool and 73 percent of the participants in the original natural frequency tree method experiment said that they plan to use this method as an analytical tool. This finding is important because it indicates that the intelligence community should seriously consider teaching and using Bayesian reasoning. As shown by this experiment, a Bayesian reasoning method can easily be taught to intelligence analysts and be applied to intelligence problems. Since the participants in this study are considered to be representative of intelligence analysts already within the intelligence community, it can thus be assumed that intelligence analysts in the intelligence community would be interested in learning and applying a method for Bayesian reasoning.

Staying in the educational vein, another important conclusion is that changes made to the natural frequency tree method can lead to improved results. While statistical analysis does not indicate that the changes made to the method were significant, anecdotal evidence indicates that the changes made to the frequency tree method did lead to better performance and it is likely that with more changes to the methodology and training session significant improvement in performance could be obtained. When compared to the participants in the original experiment, the participants in the revised natural frequency tree experiment who understood the method seemed to really understand the method and were able to obtain the correct answer faster and more accurately.

The rest of this chapter will focus on the idea of obtaining better performance and understanding through the improvement of the paper and pencil natural frequency tree method by recommending changes for possible further research and study into this topic.

In addition, it will highlight the areas that appeared to be difficult for the participants. Feedback from participants on the post-test questionnaire indicated that they felt more practice problems would have helped them understand the method better. While the revised natural frequency tree method experiment did add one extra practice problem, one more practice problem was not enough. During the revised natural frequency tree method experiment, I walked through one practice problem with the participants and then had them do one on their own. Future investigators building on this study should begin by walking through at least two practice problems before having the participants or students do a problem on their own. The natural frequency tree method allows for a lot of movement and having multiple examples will help students become familiar with how to move around the natural frequency tree and see how the different nodes that make up the natural frequency tree are related. Additionally, the more participation, the better future results will be. While students were asked to help fill in the example natural frequency tree, many did not participate. Participation is very important because it encourages questions and will point out the areas that confuse people the most. One of these areas was identifying where all of the information should be placed in the natural frequency tree.

In addition to concluding this from the results of the test questions, many of the participants indicated on their questionnaires that they had difficulty understanding exactly where the pieces of information from the problem fit into the natural frequency tree. Additional practice problems and encouraged participation would also address this difficulty. Having the students tell the instructor where the information fits into the natural frequency tree helps the instructor ensure that students are able to identify the different parts of the problem.

Another suggestion the participants provided in the post-test questionnaire feedback was to have a follow-up session. This is a very good idea as it will allow the students to take time to review the information learned and practice on their own without being under the pressure to perform well on the spot. Instead of having a follow-up session that only tests for the transferability of knowledge like Sedlmeier and Gigerenzer, the follow-up session should also be an opportunity for students to ask questions to the instructor on a one-by-one basis as well as do some more practice problems before being tested. Having a follow-up session like this will likely encourage students to bring questions they might not have thought of during the initial training session and also ones that they may not have felt comfortable asking in front of everyone before. Once the students feel comfortable with the method they should be asked to answer the test questions.

The main problem with having a follow-up session is retention. Unless the initial training session and follow-up session are part of a classroom exercise, it may be difficult to bring enough students back in order to test their abilities. Sedlmeier and Gigerenzer ran into this same problem. A solution that they came up with and one that could be applied in this case is to only offer the reward, i.e., extra credit if doing the experiment outside of the classroom, to those who show up for both sessions. This change in methodology has an excellent chance of improving the performance results of intelligence analysts who use the paper and pencil natural frequency tree method.

A final recommendation for improvement is to provide the natural frequency tree method procedures to the students ahead of time. This recommendation, however, would only likely be significantly effective if the experiment was being used as a classroom exercise since it could then be assigned by the professor as homework. It is unlikely that students would read the procedures ahead of time if the experiment was taking place outside of the classroom. Reading the procedures ahead of time would allow students to familiarize themselves with the paper and pencil natural frequency tree method and be able to come prepared to the training session as well as bring questions. From observation during the natural frequency tree experiments, it appeared that very few of the participants actually read the procedures. Most participants paid attention to the presentation and then tried to do the problems on their own. While they did end up referring back to the procedures handout, if the participants had read the procedures ahead of time they would have been able to easily refer to the different sections and have a better understanding at the same time.

The qualitative feedback from the participants offered many insights into modifications that could be made to the natural frequency tree method. These changes are likely to lead to improved performance and understanding of the method. It appears from the limited results found in this experiment that the paper and pencil natural frequency tree method with improvements has some potential to be a useful analytic tool in the intelligence community.

In addition to making changes to the paper and pencil natural frequency tree method, a new direction of study could look at developing a similar type of method that involved multiple pieces of evidence and/or additional hypotheses. The current natural frequency tree method only takes into account one piece of evidence. An analyst in the intelligence field will be faced with numerous pieces of information and will be required to evaluate the impact of that information on their hypotheses. A method that incorporates this aspect of analysis would be very useful to the intelligence analyst. Additionally, a method that would allow the analyst to add additional hypotheses would also be very useful.

Bayesian reasoning offers several benefits to the intelligence community. This experiment showed that it is possible to teach intelligence analysts a new, structured method in a small amount of time. In addition, this study showed qualitative findings that analysts, especially those who will be entering the intelligence community soon, are interested in learning new structured analytical techniques that will help them make better estimates and reduce uncertainty. This thesis only provides a starting point for the vast amount of research that can be done on the benefits of using Bayesian reasoning in the intelligence field.

BIBLIOGRAPHY

Ajzen, Icek. "Intuitive Theories of Events and the Effects of Base-Rate Information on Prediction." *Journal of Personality and Social Psychology* 35, no. 5 (1977): 303-314.

Bar-Hillel, Maya. "The Base-Rate Fallacy In Probability Judgments." *Acta Psycholigica* 44, (1980): 211-233.

Blair, Bruce. "The Logic of Intelligence Failure." Center for Defense Information, March 9, 2004, http://www.cdi.org/blair/logic.cfm (accessed Apr. 2, 2007).

Casscells,W., A. Schoenberger, and T.B. Graboys. "Interpretation by physicians of clinical laboratory results." *New England Journal of Medicine* 299 (1978): 999-1001.

Cosmides, Leda, and John Tooby. "Are humans good intuitive statisticians after all? Rethinking some conclusions form the literature on judgment under uncertainty." *Cognition* 58 (1996): 1-73.

Eddy, David M. "Probabilistic reasoning in clinical medicine: Problems and opportunities." In *Judgment under uncertainty: Heuristics and biases*, eds. Daniel Kahneman, Paul Slovic, and Amos Tversky, 249-267. Cambridge: Cambridge University Press, 1982.

Fisk, Charles E. "The Sino-Soviet Border Dispute: A Comparison of the Conventional and Bayesian Methods for Intelligence Warning." *Studies in Intelligence* 16, no. 2 (1972): 52-62. https://www.cia.gov/csi/kent_csi/pdf/v16i2a04p.pdf#page=1. (accessed Mar. 22, 2007).

Folker, Jr., MSgt Robert D. (USAF). "Intelligence Analysis in Theater Joint Intelligence Centers: An Experiment in Applying Structured Methods." Thesis, Joint Military Intelligence College, 2000.

Gigerenzer, Gerd. "The Psychology of Good Judgment: Frequency Formats and Simple Algorithms." *Medical Decision Making* 19 (1996): 273-280.

Gigerenzer, Gerd, and Adrian Edwards. "Simple tools for understanding risks: from innumeracy to insight." *BMJ* 327, (2003): 741-744. http://www.bmj.com/cgi/reprint/327/7417/741 (accessed Mar.15, 2007).

Gigerenzer, Gerd, and Ulrich Hoffrage. "How to Improve Bayesian Reasoning Without Instruction: Frequency Formats." *Psychological Review* 102, no. 4 (1995): 684-704.

Gigerenzer, Gerd, Zeno Swijtink, Theodore Porter, Lorraine Daston, John Beatty, and Lorenz Krüger. *The Empire of Chance: How probability changed science and everyday life*. Cambridge: Cambridge University Press, 1989.

Ginosar, Zvi, and Yaacov Trope. "The Effects of Base Rates and Individuating Information on Judgments about Another Person." *Journal of Experimental Social Psychology* 16 (1980): 228-242.

Heuer, Jr., Richards J. "Limits of Intelligence Analysis." *Orbis* (2005): 75-94.

Heuer, Jr., Richards J. *Psychology of Intelligence Analysis*. Langley, Virginia: Central Intelligence Agency, 1999.

Heuer, Jr. Richards J., ed. *Quantitative Approaches to Political Intelligence: The CIA Experience*. Boulder, Colorado: Westview Press, Inc., 1978.

Hoffrage, Ulrich and Gerd Gigerenzer. "Using Natural Frequencies to Improve Diagnostic Inferences." *Academic Medicine* 73 (1998): 538-540.

Howson, Colin, and Peter Urbach. *Scientific Reasoning: The Bayesian Approach*. La Salle, Illinois: Open Court Publishing Company, 1989.

Hutchinson, John M.C., and Gerd Gigerenzer. "Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet." *Behavioural Processes* 69 (2005): 97-124.

Iversen, Gudmund R. *Bayesian Statistical Inference*. Beverly Hills: Sage Publications, Inc., 1984.

Johnson, Janet Buttolph, and Richard A. Joslyn. *Political Science Research Methods*. Washington, DC: CQ Press, 1986.

Johnston, Rob. *Analytic Culture in the U.S. Intelligence Community: An Ethnographic Study*. Washington, DC: Center for the Study of Intelligence, 2005.

Kahneman, Daniel, and Amos Tversky. "On The Psychology Of Prediction." *Psychological Review* 80, no. 4 (1973): 237-251.

Kahneman, Daniel, and Amos Tversky. "Subjective Probability: a Judgment of Representativeness." *Cognitive Psychology* 3, no. 3 (1972): 430-454.

Koehler, Jonathan J. "The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges." *Behavioral and Brain Sciences* 19 (1996): 1-53.

Krämer, Walter, and Gerd Gigerenzer. "How to Confuse with Statistics or: The Use and Misuse of Conditional Probabilities." *Statistical Science* 20, no. 3 (2005): 223-230.

Lee, Peter M. *Bayesian Statistics: An Introduction*. London: Arnold Publishers, 1997.

Lichtenstein, Sarah, Baruch Fischhoff, and Lawrence D. Phillips. "Calibration of probabilities: The state of the art to 1980." In *Judgment under uncertainty: Heuristics and biases*, eds. Daniel Kahneman, Paul Slovic, and Amos Tversky, 306-334. Cambridge: Cambridge University Press, 1982.

Malakoff, David. "Bayes Offers a 'New Way' to Make Sense of Numbers." *Science* 286 (1999): 1460-1464.

Martignon, Laura and Christoph Wassner. "Teaching Decision Making and Statistical Thinking with Natural Frequencies." Paper presented at the International Conference on the Teaching of Statistics 6, South Africa, 2002.

McLaughlin, Jessica. "A Bayesian Updating Model for Intelligence Analysis: A Case Study of Iraq's Nuclear Weapons Program." Thesis, Stanford University, 2005.

Nisbett, Richard E., and Eugene Borgida. "Attribution and the Psychology of Prediction." *Journal of Personality and Social Psychology* 32, no. 5 (1975): 932-943.

O'Connor, J.J., and E.F. Robertson. "*Pierre-Simon Laplace*." Jan. 1999. http://www-groups.dcs.st-andrews.ac.uk/%7Ehistory/Mathematicians/Laplace.html (accessed Mar. 22, 2007).

O'Connor, J.J., and E.F. Robertson. "*Thomas Bayes*." June 2004. http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Bayes.html (accessed Mar. 22, 2007).

Pollard, Paul, and Jonathan St. B.T. Evans. "The role of 'representativeness' in statistical inference: a critical appraisal." In *Thinking and reasoning: psychological approaches*, ed. Jonathan St. B.T. Evans, 107-134. London: Routledge & Kegan Paul, 1983.

Reis, Harry T., and Charles M. Judd, eds. *Handbook Of Research Methods In Social And Personality Psychology.* Cambridge: Cambridge University Press, 2000.

Schrage, Michael. "What Percent is 'Slam Dunk'?." *Washington Post*, February 20, 2005, Page B01, http://www.washingtonpost.com/wp-dyn/articles/A37115-2005Feb19.html (accessed Apr. 2, 2007).

Schweitzer, Nicholas. "Bayesian Analysis for Intelligence: Some Focus on the Middle East." *Studies in Intelligence* 20, no. 2 (1976): 31-44. https://www.cia.gov/csi/kent_csi/pdf/v20i2a03p.pdf#page=1 (accessed Mar. 22, 2007).

Sedlmeier, Peter and Gerd Gigerenzer. "Teaching Bayesian Reasoning in Less Than Two Hours." *Journal of Experimental Psychology: General* 130, no. 3 (2001): 380-400.

Slovic, Paul, Baruch Fischhoff, and Sarah Lichtenstein. "Behavioral Decision Theory." *Annual Review of Psychology* 28 (1977): 1-39.

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

Tversky, Amos, and Daniel Kahneman. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (1974): 1124-31.

Tversky, Amos, and Daniel Kahneman. "Causal Schemas in Judgments Under Uncertainty." In *Progress in Social Psychology* Vol. 1, ed. Martin Fishbein, 49-72. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers, 1980.

Westbury, Chris. "Bayes' For Beginners." Paper written for author's 4th year Psychometrics class, University of Alberta, 2005.

Wheaton, Kristan J., Jennifer Lee, and Hemangini Deshmukh. "Bayesian Statistics In The Real World Of Intelligence Analysis: Lessons Learned." Paper presented at the International Studies Association's 2007 Annual Convention, Chicago, Illinois, February 28 – March 3, 2007.

United States Government. *Intelligence Reform and Terrorism Prevention Act of 2004*. December 17, 2004. http://www.nctc.gov/docs/pl108_458.pdf (accessed Mar. 22, 2007).

United States Government. *Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction*. March 31, 2005. http://www.wmd.gov/report/ (accessed Apr. 8, 2007).

United States Government. *The 9/11 Commission Report*. July 22, 2004. http://www.gpoaccess.gov/911/pdf/fullreport.pdf (accessed Apr. 8, 2007).

Wikipedia. "List of Cognitive Biases." http://en.wikipedia.org/wiki/Cognitive_biases (accessed Apr. 24, 2007).

Yudkowsky, Eliezer. "An Intuitive Explanation of Bayesian Reasoning." http://yudkowsky.net/bayes/bayes.html (accessed Mar. 26, 2007).

Zlotnick, Jack. "Bayes' Theorem For Intelligence Analysis." *Studies in Intelligence* 16, no. 2 (1972): 43-52. https://www.cia.gov/csi/kent_csi/pdf/v16i2a03p.pdf#page=1 (accessed Mar. 21, 2007).

Zlotnick, Jack. “A Theorem for Prediction.” *Studies in Intelligence* 11, no. 4 (1967): 1-12. https://www.cia.gov/csi/kent_csi/pdf/v11i4a01p.pdf#page=1 (accessed Mar. 22, 2007).

APPENDICES

APPENDIX A

List of Cognitive Biases from Wikipedia[149]

Decision-making and behavioral biases

Many of these biases are studied for how they affect belief formation and business decisions and scientific research.

- Bandwagon effect — the tendency to do (or believe) things because many other people do (or believe) the same. Related to groupthink, herd behavior, and manias.

- Bias blind spot — the tendency not to compensate for one's own cognitive biases.

- Choice-supportive bias — the tendency to remember one's choices as better than they actually were.

- Confirmation bias — the tendency to search for or interpret information in a way that confirms one's preconceptions.

- Congruence bias — the tendency to test hypotheses exclusively through direct testing, in contrast to tests of possible alternative hypotheses.

- Contrast effect — the enhancement or diminishment of a weight or other measurement when compared with recently observed contrasting object.

- Déformation professionnelle — the tendency to look at things according to the conventions of one's own profession, forgetting any broader point of view.

- Endowment effect — "the fact that people often demand much more to give up an object than they would be willing to pay to acquire it."

- Focusing effect — prediction bias occurring when people place too much importance on one aspect of an event; causes error in accurately predicting the utility of a future outcome.

- Hyperbolic discounting — the tendency for people to have a stronger preference for more immediate payoffs relative to later payoffs, the closer to the present both payoffs are.

- Illusion of control — the tendency for human beings to believe they can control or at least influence outcomes that they clearly cannot.

[149] Wikipedia, "List of Cognitive Biases," http://en.wikipedia.org/wiki/Cognitive_biases.

- Impact bias — the tendency for people to overestimate the length or the intensity of the impact of future feeling states.
- Information bias — the tendency to seek information even when it cannot affect action.
- Loss aversion — "the disutility of giving up an object is greater than the utility associated with acquiring it."
- Neglect of probability — the tendency to completely disregard probability when making a decision under uncertainty.
- Mere exposure effect — the tendency for people to express undue liking for things merely because they are familiar with them.
- Omission bias — the tendency to judge harmful actions as worse, or less moral, than equally harmful omissions (inactions).
- Outcome bias — the tendency to judge a decision by its eventual outcome instead of based on the quality of the decision at the time it was made.
- Planning fallacy — the tendency to underestimate task-completion times.
- Post-purchase rationalization — the tendency to persuade oneself through rational argument that a purchase was a good value.
- Pseudocertainty effect — the tendency to make risk-averse choices if the expected outcome is positive, but make risk-seeking choices to avoid negative outcomes.
- Selective perception — the tendency for expectations to affect perception.
- Status quo bias — the tendency for people to like things to stay relatively the same.
- Von Restorff effect — the tendency for an item that "stands out like a sore thumb" to be more likely to be remembered than other items.
- Zero-risk bias — preference for reducing a small risk to zero over a greater reduction in a larger risk.

Biases in probability and belief

Many of these biases are often studied for how they affect business and economic decisions and how they affect experimental research.

- Ambiguity effect — the avoidance of options for which missing information makes the probability seem “unknown.”

- Anchoring — the tendency to rely too heavily, or “anchor,” on one trait or piece of information when making decisions.

- Anthropic bias — the tendency for one's evidence to be biased by observation selection effects.

- Attentional bias — neglect of relevant data when making judgments of a correlation or association.

- Availability heuristic — a biased prediction, due to the tendency to focus on the most salient and emotionally-charged outcome.

- Clustering illusion — the tendency to see patterns where actually none exist.

- Conjunction fallacy — the tendency to assume that specific conditions are more probable than general ones.

- Gambler's fallacy — the tendency to assume that individual random events are influenced by previous random events. For example, “I've flipped heads with this coin so many times that tails is bound to come up sooner or later.”

- Hindsight bias — sometimes called the “I-knew-it-all-along” effect, the inclination to see past events as being predictable.

- Illusory correlation — beliefs that inaccurately suppose a relationship between a certain type of action and an effect.

- Ludic fallacy — the analysis of chance related problems with the narrow frame of games. Ignoring the complexity of reality, and the non-gaussian distribution of many things.

- Neglect of prior base rates effect — the tendency to fail to incorporate prior known probabilities which are pertinent to the decision at hand.

- Observer-expectancy effect — when a researcher expects a given result and therefore unconsciously manipulates an experiment or misinterprets data in order to find it.

- Optimism bias — the systematic tendency to be over-optimistic about the outcome of planned actions.

- Overconfidence effect — the tendency to overestimate one's own abilities.

- Positive outcome bias — a tendency in prediction to overestimate the probability of good things happening to them.

- Recency effect — the tendency to weigh recent events more than earlier events.

- Reminiscence bump — the effect that people tend to recall more personal events from adolescence and early adulthood than from other lifetime periods.

- Rosy retrospection — the tendency to rate past events more positively than they had actually rated them when the event occurred.

- Primacy effect — the tendency to weigh initial events more than subsequent events.

- Subadditivity effect — the tendency to judge probability of the whole to be less than the probabilities of the parts.

- Telescoping effect — the effect that recent events appear to have occurred more remotely and remote events appear to have occurred more recently.

- Texas sharpshooter fallacy — the fallacy of selecting or adjusting a hypothesis after the data are collected, making it impossible to test the hypothesis fairly.

Social biases

Most of these biases are labeled as attributional biases.

- Actor-observer bias — the tendency for explanations for other individual's behaviors to overemphasize the influence of their personality and underemphasize the influence of their situation. This is coupled with the opposite tendency for the self in that one's explanations for their own behaviors overemphasize their situation and underemphasize the influence of their personality.

- Egocentric bias — occurs when people claim more responsibility for themselves for the results of a joint action than an outside observer would.

- Forer effect (aka Barnum Effect) — the tendency to give high accuracy ratings to descriptions of their personality that supposedly are tailored specifically for them, but are in fact vague and general enough to apply to a wide range of people. For example, horoscopes.

- False consensus effect — the tendency for people to overestimate the degree to which others agree with them.

- Fundamental attribution error — the tendency for people to over-emphasize personality-based explanations for behaviors observed in others while under-emphasizing the role and power of situational influences on the same behavior.

- Halo effect — the tendency for a person's positive or negative traits to “spill over” from one area of their personality to another in others' perceptions of them.

- Illusion of asymmetric insight — people perceive their knowledge of their peers to surpass their peers’ knowledge of them.

- Illusion of transparency — people overestimate others’ ability to know them, and they also overestimate their ability to know others.

- In group bias — preferential treatment people give to whom they perceive to be members of their own groups.

- Just-world phenomenon — the tendency for people to believe that the world is “just” and therefore people ‘get what they deserve.”

- Lake Wobegon effect — the human tendency to report flattering beliefs about oneself and believe that one is above average.

- Notational bias — a form of cultural bias in which a notation induces the appearance of a nonexistent natural law.

- Out group homogeneity bias — individuals see members of their own group as being relatively more varied than members of other groups.

- Projection bias — the tendency to unconsciously assume that others share the same or similar thoughts, beliefs, values, or positions.

- Self-serving bias — the tendency to claim more responsibility for successes than failures. It may also manifest itself as a tendency for people to evaluate ambiguous information in a way beneficial to their interests.

- Self-fulfilling prophecy — the tendency to engage in behaviors that elicit results which will (consciously or subconsciously) confirm our beliefs.

- System justification — the tendency to defend and bolster the status quo, i.e. existing social, economic, and political arrangements tend to be preferred, and alternatives disparaged sometimes even at the expense of individual and collective self-interest.

- Trait ascription bias — the tendency for people to view themselves as relatively variable in terms of personality, behavior and mood while viewing others as much more predictable.

APPENDIX B

Question Using Standard Statistical Terminology Used in the Natural Frequency Verification Experiment

The probability that Country X will go to war in any given month with Country Y is 1%

We receive a new SIGINT report that suggests that Country X is about to go to war with Country Y. We assess the likelihood that this report is accurate at about 95%. We assess the odds that it is deliberate deception to be about 5%.

What are the **approximate odds** that Country X is actually going to go to war?

APPENDIX C

Question Using Natural Frequencies Used in the Natural Frequency Verification Experiment

The Intelligence Community of Country X has about 100,000 people working for it. About 100 of them are spies.

All 100,000 of the employees in the Intelligence Community have to take a special test. The test will positively identify 95 of the 100 spies that take the test as spies.

The test will also positively identify 50 of every 1000 people that take the test as spies even though they are not spies.

Your best friend has just taken the special test and was positively identified as a spy. What are the **approximate odds** that your best friend is actually a spy?

APPENDIX D

Mercyhurst College Institutional Review Board Submission and Approval Letter

Submission to the Institutional Review Board

Date Submitted:
January 4, 2007

Investigator(s):
Jennifer Lee

Investigator Address:
Street Address
City, State Zip

Investigator(s) E-mail:
XXX

Investigator Telephone Number:
XXX-XXX-XXXX

Advisor's Name (if applicable):
Kristan Wheaton

Advisor's E-mail:
XXX

Advisor's Signature of Approval:
[*X*] Place X here if advisor has approved research

Title of Research Project:
Alternative Bayesian Reasoning Method For The Intelligence Community

Date of Initial Data Collection:
TBD, anticipate late January

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

Bayesian reasoning is a statistical technique that combines base rates with evidence to help decision makers make accurate decisions. Unfortunately, for most people, Bayesian reasoning is a complicated technique to learn. In Sedlmeier's and Gigerenzer's (2001) study on Bayesian reasoning, they proposed alternative, yet mathematically equal, techniques to perform Bayesian reasoning. According to the authors, when Bayesian problems are presented in terms of natural frequencies, terms that provide information regarding base rates, it is easier for people without strong statistical backgrounds to use Bayesian reasoning to answer conditional probability questions. To test their alternative techniques, Sedlmeier and Gigerenzer used a computer training program that taught participants how to perform their frequency grid and frequency tree techniques. The authors' results showed that by using these techniques, people were able to learn and perform Bayesian reasoning.

Bayesian reasoning has long been discussed as a possible analytical tool that would be useful for the intelligence community. I have developed a paper and pencil method based on Sedlmeier's and Gigerenzer's computerized frequency tree technique. I plan to test my method using intelligence problems and intelligence analysts (both undergraduate *and graduate students in the intelligence program) to determine if my method is a viable analytical technique that intelligence analysts can use for Bayesian reasoning.*

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

Materials:

PowerPoint (to teach my frequency tree method)
Test Problems
Post-Test Questionnaire

Procedure:

One week prior to testing my frequency tree method, I will send the participants my frequency tree method instructions (attached at end) asking them to review the instructions before the training session. At the training session, I will use PowerPoint to present the method and answer any questions. Once the training is complete and there are no other questions, I will ask the participants to apply my method to five different intelligence-related Bayesian problems (attached at end). The problems will vary in layout, thus allowing me to test my method for its applicability to a variety of Bayesian problems. Following the completion of the test problems, I will give the participants the answers to the problems and ask them to fill-out a questionnaire (attached at end) and provide feedback regarding both the method and the training session.

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 30 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: *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 (it is a good educational experience after all).*

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 session, 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 will be happening. Following the administration of the questionnaire, participants will be provided with a debriefing statement that will explain how the results from the session will be used.

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 Dr. Terry F. Pettijohn; Institutional Review Board Chair; Mercyhurst College; 501 East 38th Street; Erie, Pennsylvania 16546-0001; Telephone (XXX) XXX-XXXX."

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:
Bayesian reasoning is an analytical tool that can benefit any intelligence analyst. By gaining an understanding of Bayesian reasoning, participants will be able to add an analytical technique, which is not taught in the intelligence program, to the collection of techniques they have already gained from previous coursework.

For the Intelligence Community:
Bayesian reasoning allows decision makers to make more accurate decisions because it takes into consideration both the possibility of an event happening and an event not happening. A better understanding of Bayesian reasoning may benefit the United States government agencies and anyone in a position that requires making decisions.

Please submit this file and accompanying materials to the IRB Chair, Terry Pettijohn, via electronic mail (XXX) for review. *(revised 8/2006 tfp)*

Institutional Review Board Approval Letter

Jennifer Lee
Mercyhurst College

January 14, 2007

Ms. Lee,

The Mercyhurst College Institutional Review Board has reviewed your research proposal entitled “Alternative Bayesian Reasoning Method For The Intelligence Community.” Based on committee member input, and your recent revisions, your proposal has been approved. You may now begin your project.

If you should run into any significant difficulties involving the ethical treatment of research participants in conducting your study, or if you wish to make significant changes in your procedure, please inform the IRB.

If you have any questions or require any additional information, please contact me by e-mail (XXX) or phone (XXX-XXX-XXXX).

Good luck with your research.

Sincerely,

Terry F. Pettijohn II, Ph.D.
Chair, Institutional Review Board

APPENDIX E

Email Sent to All Students in the Intelligence Studies Program at Mercyhurst College for the First Natural Frequency Tree Method Experiment

To All Intelligence Studies Students,

Jennifer Lee, a second year graduate student in the intelligence program, has developed a Bayesian Reasoning method for her thesis and is looking for participants to take the time to learn her method and provide feedback regarding this analytical technique.
Bayesian Reasoning is a great technique to understand. It will make you a better analyst and give you a competitive edge when it comes to talking to employers. Participation involves learning her method to perform Bayesian Reasoning, applying the method to a few problems, and filling out a questionnaire.

- You **do not need** to have a strong mathematical background to perform this analytical technique.

- The total time commitment required for your participation is **1 ½ hours**.

- Most professors have agreed to offer **extra credit** for your participation. Check with your individual professors for more detail about what they will offer in terms of extra credit.

This is your chance to gain an understanding of Bayesian Reasoning. If you are interested in participating, **please let Jen know via email at XXX** at least 24 hours prior to the instructional session, which will be held on Tuesday, January 30 at 5:00 PM in Zurn 214.

Please **do not** respond to the RIAP group.

APPENDIX F

Participation Consent Form

Bayesian Reasoning
Participation Consent Form

The purpose of this research is to test a method for Bayesian reasoning and determine if it is a viable analytical tool for intelligence analysts.

Your participation involves learning a new method for Bayesian reasoning, solving post-training intelligence-related problems, and filling out a questionnaire. This process should take no longer than 1.5 hours. 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 Bayesian reasoning or this research you can contact me at XXX.

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 Dr. Terry F. Pettijohn; Institutional Review Board Chair; Mercyhurst College; 501 East 38th Street; Erie, Pennsylvania 16546-0001; Telephone (XXX) XXX-XXXX.

Jennifer Lee, *Applied Intelligence Master's Student, Mercyhurst College* XXX-XXX-XXXX
Kristan Wheaton, *Research Advisor, Mercyhurst College* XXX-XXX-XXXX

APPENDIX G

Frequency Tree Method Procedures

Frequency Tree Method Instructions

By: Jennifer Lee

Tools needed:
Paper
Pencil or Pen
Calculator

Training Case:

This example is comprised solely of ***fictitious*** information and is for training purposes only.

According to a recent analytical report, you know there is a 2 percent chance Japan is creating nuclear weapons. While sifting through your daily traffic, you come across a SIGINT report indicating Japan is stockpiling plutonium for creating nuclear weapons. From previous experience, you know that a SIGINT report from this source is 95 percent accurate. However, being the analyst you are, you assess there is a 3 percent chance that it is deliberate deception.

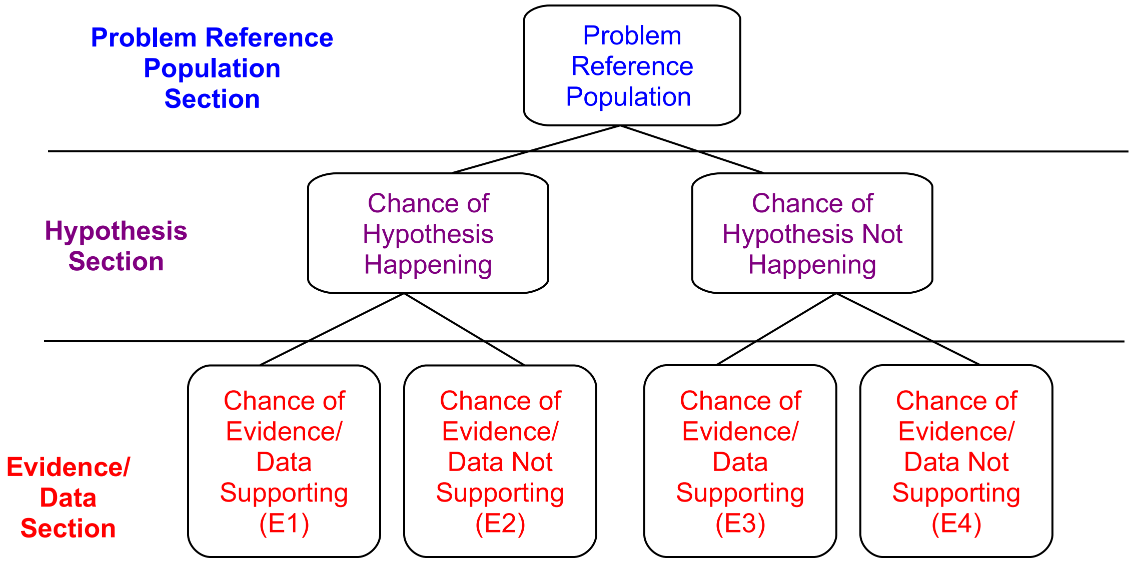
What are the approximate odds that Japan is actually stockpiling plutonium for the creation of nuclear weapons?

Frequency Tree Steps:

1. Draw Frequency Tree
2. Create Natural Frequencies And Fill In Data
3. Calculate Answer

Step 1: Draw Frequency Tree

The frequency tree is comprised of three sections (see next page):



The Problem Reference Population section will most likely not be given in the problem data. This section will be calculated from the information in the Hypothesis section.

The Hypothesis section includes the information from the problem that describes the chances of the main event or action happening or not happening. This is usually what your new evidence suggests.

The Evidence/Data section includes the information that acts as evidence of the event or action happening or not happening. The evidence can either support the chances of the main event or action happening or it can support the chances of the main event or action not happening.

For each problem, draw an empty frequency tree. Step 2 will discuss how to fill in the frequency tree using the data from the problem.

Check List Before Moving To Next Step:

☑ Draw empty frequency tree.

Step 2: Create Natural Frequencies And Fill In Data

Natural frequencies, simply defined, put percentages in terms that are easily understandable to those who do not have strong statistical backgrounds. What natural frequencies provide are the reference populations to the percentages. For example, 10 percent can easily be put into a natural frequency by saying “10 out of 100”. “10 out of 100” is equal to 10 percent and more easily understood. In addition, “10 out of 100” provides information about the reference population, which is 100 in this example.

Step 2A:

Before creating the natural frequencies, determine where the information you currently have from the problem fits into the frequency tree. It is also very important to identify at this time what the hypothesis is, or the question you are trying to answer. This is the ***most important*** step, as it identifies where in the frequency tree each piece of goes. The best thing to do is to write out next to your frequency tree what each piece of information represents, as shown below. Keep in mind that each section is the reference population for the section below it. Below are the pieces of information from the problem above. They have been listed and labeled as to where they fit into the frequency tree:

Hypothesis: Japan is stockpiling plutonium for the creation of nuclear weapons.

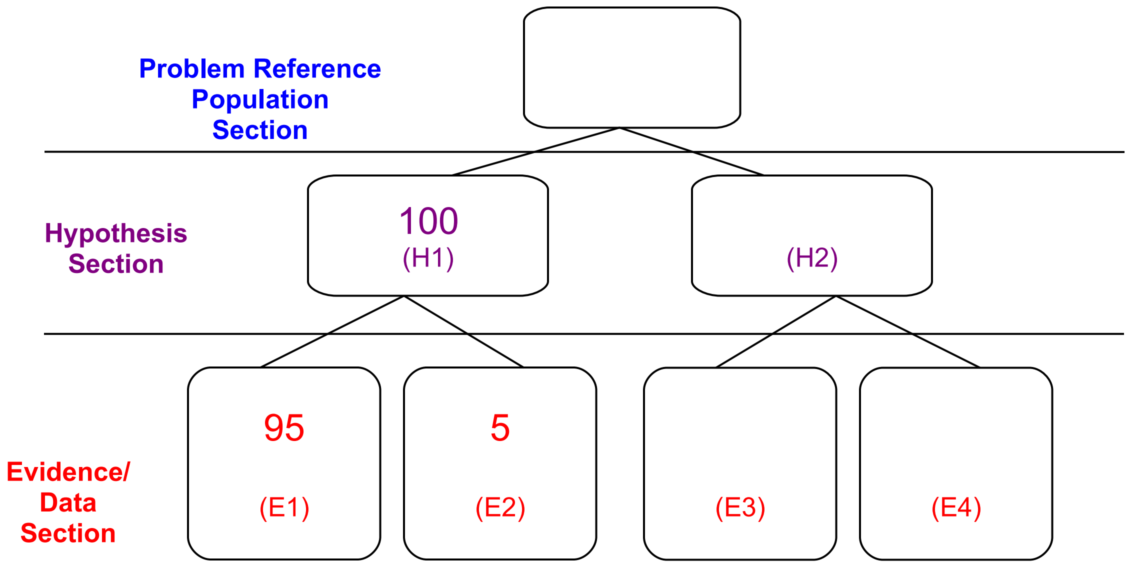
Hypothesis Section (H1): There is a 2 percent chance that Japan is stockpiling plutonium for the creation of nuclear weapons.

Evidence/Data Section (E1): A SIGINT report indicates Japan is creating nuclear weapons. From previous experience, you know that a SIGINT report from this source is 95 percent accurate.

Evidence/Data Section (E3): You assess there is a 3 percent chance that it is deliberate deception.

To put statistical information into natural frequencies easily, skim the problem for the largest given percentage. In the example above, that would be 95. To keep 95 a whole number, its reference population needs to be 100. The reference population for the first number will always be 100. The 95 goes into the box E1 in the Evidence/Data section (as described above), and since the level above it is its reference population, the 100 goes into the Hypothesis section’s H1 box.

Keep in mind that the largest number ***will not*** necessarily always go in box E1 in the Evidence/Data section. Depending on the problem, the largest number could go in any box in the Hypothesis section or Evidence/Data section. Another key point to remember is that the boxes in the section below the reference population always need to add up to the reference population value. In this problem, since boxes E1 and E2 underneath box H1 need to add up to the reference population, which in this problem is 100, the value that goes into box E2 in the Evidence/Data section is 5.



Check List Before Moving To Next Step:

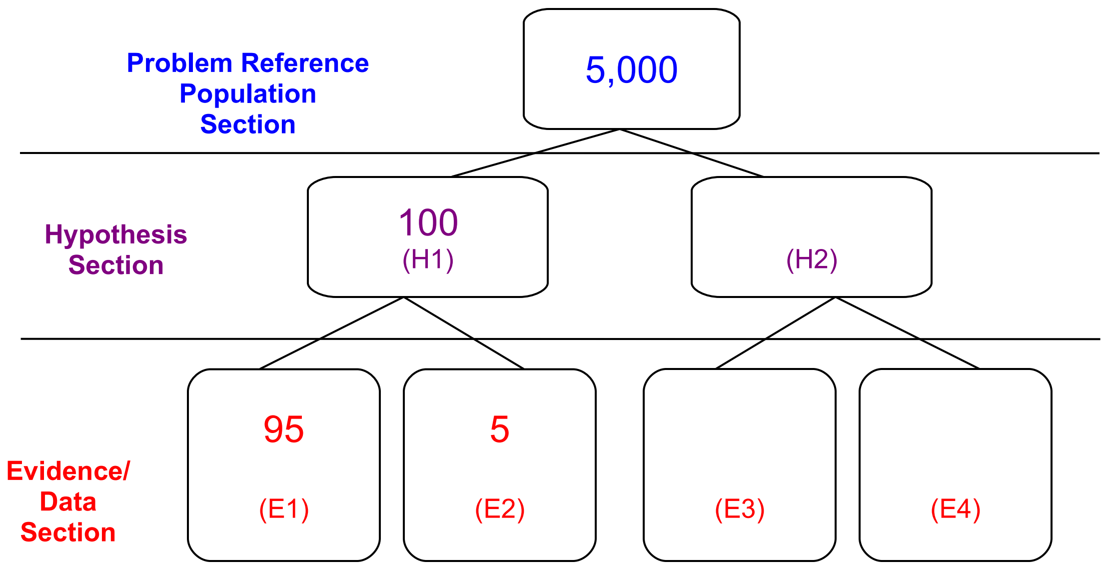
- Write down each piece of information from the problem and where each piece fits into frequency tree.
- Put the largest given number into the correct box in the frequency tree and made the value (reference population) in the box above it 100 according to the information given in the problem.

Step 2B:

Once the first piece of information is filled-in, the next step is to determine the reference population for the box above the filled-in boxes. In the above example, this would be determining the reference population for the 100 in the Hypothesis section. From the information in the problem, we know that the 100 is equal to 2 percent of the Problem Reference Population. To determine the Problem Reference Population, divide the Hypothesis section reference population in box

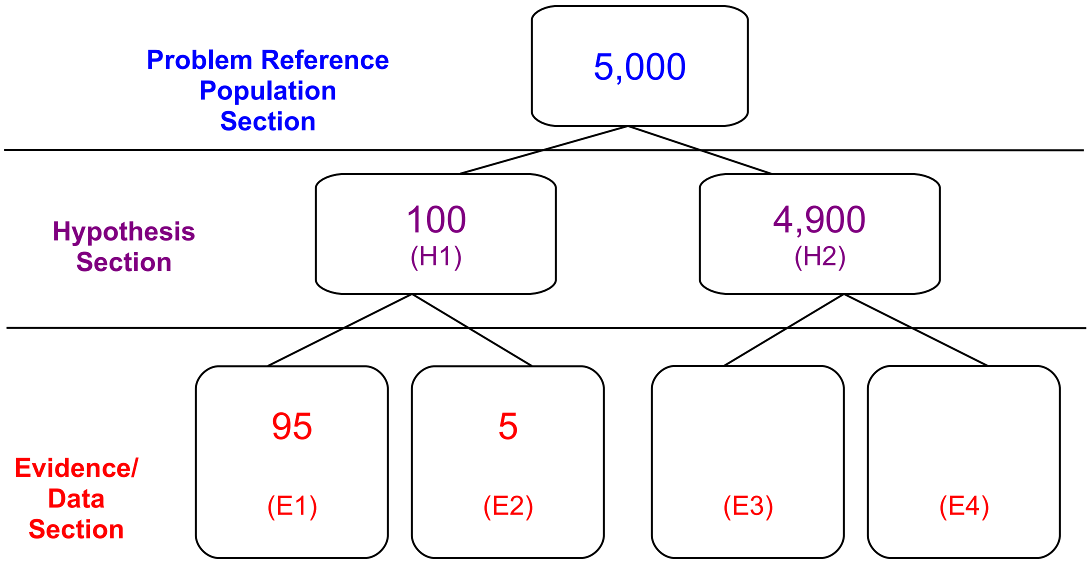
H1, which is the 100, by 2 percent, or 0.02. Doing this simple calculation provides the Problem Reference Population, which is 5,000.

It is important to remember that it does not matter what the reference population turns out to be as long as the ratios between the reference population and the boxes below it match. It is ***very important*** to keep the ratios correct. In this example, the reference population is 5,000, but in another problem, depending on the ratios, the reference population could be 2,000. Additionally, all of the reference populations should be whole numbers, even if this means having to round up or down. Since you are most likely only looking for an approximate answer to the question, rounding should not be an issue.



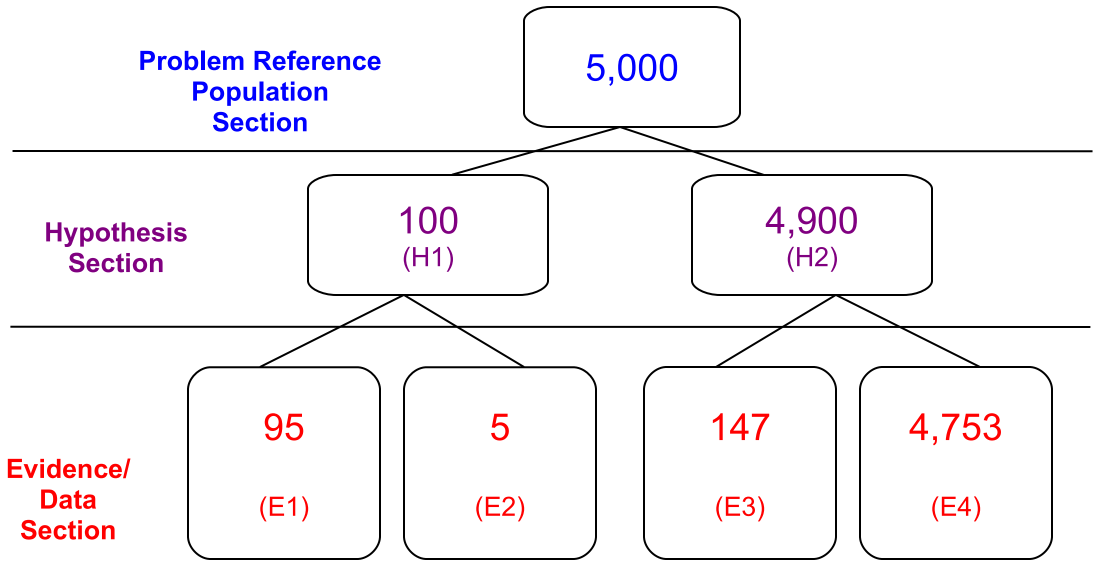
Step 2C:

The next step is to fill in box H2 in the Hypothesis section. Since we know that the two Hypothesis section boxes need to add up to 5,000, subtract the 100 in left box H1 from the 5,000 in the Problem Reference Population section. This gives you 4,900.



Step 2D:

The final step is to calculate the values for boxes E3 and E4 in the Evidence/Data section. To do this, we need to use the last piece of evidence, which states that 3 percent of the time, there is deliberate deception (also known as the chance of the evidence supporting the main event or action happening even though the main event or action will not be happening). This number goes in box E3 in the Evidence/Data section. To determine the number, multiply 4,900 by 3 percent, or 0.03. This calculation results in 147. Once again, we know that boxes E3 and E4 in the Evidence/Data section need to add up to the reference population in Hypothesis section box H2. Therefore, we need to subtract 147 from 4,900, which results in 4,753.

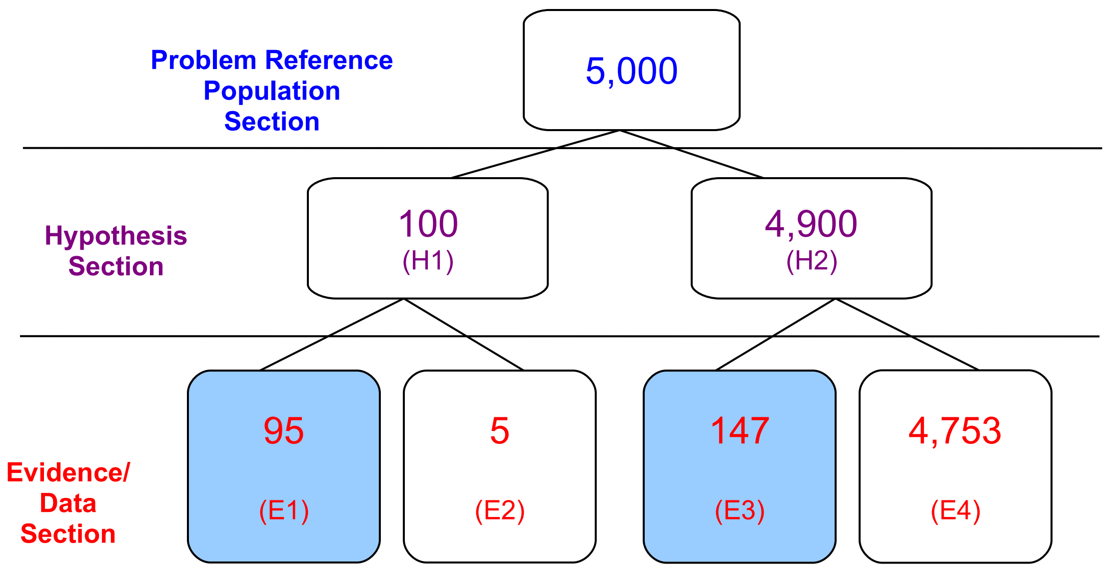


Check List Before Moving To Next Step:

- Make sure the ratios between the reference population(s) and the boxes below the reference population(s) match the information given in the problem.
- Make sure the values in the boxes below the reference population(s) add up correctly.

Step 3: Calculate Answer

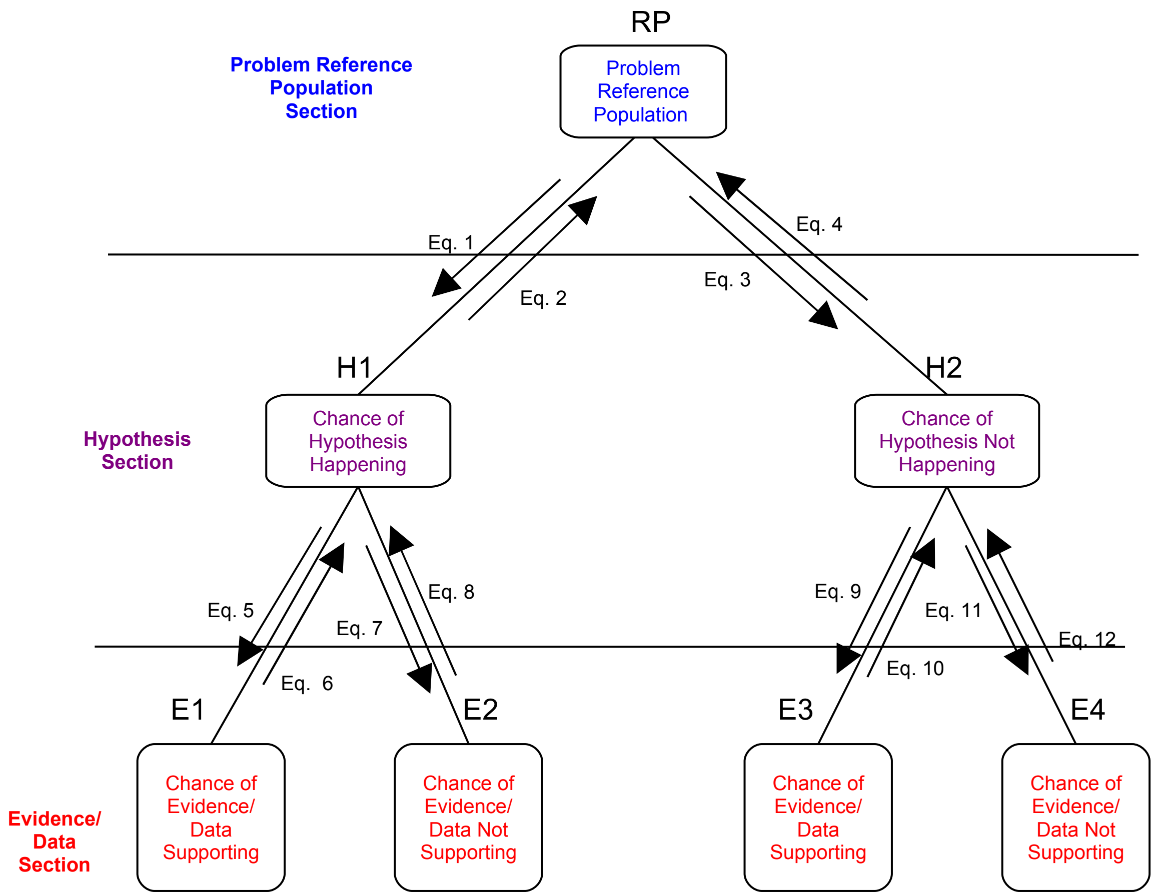
The key to calculating the answer is boxes E1 and E3 (shaded blue) in Evidence/Data section of the frequency tree.



To determine the answer to the question, "What are the approximate odds that Japan is creating nuclear weapons given this new piece of evidence?" all you need to do is divide box E1 (95) by the sum of boxes E1 and E3 (95 plus 147). Therefore, the answer to the question is 95 divided by 242 (the sum of 95 plus 147), which is .39 or 39 percent. To verify this answer, one can simply input the data into the real Bayesian formula (see appendix).

Mathematical Calculations To Move Around The Frequency Tree

As mentioned above, the largest input number (the number you initially start with) will not always be the box E1 in the Evidence/Data section. The frequency tree method allows you to input initial data into any of the Hypothesis section or Evidence/Data section boxes. The key to using the frequency tree method is correctly determining which section the given information falls into (Step 2 above). Once that has been determined, you can easily calculate the values of the boxes using the calculations below. As described in the steps above, always start with the largest number. Also, do not forget that each section's box values need to add up to the reference population in the section above it. In the diagram below, each box has a reference (that corresponds to the box references used above) above it since the calculations will refer to specific boxes using these references. The arrows next to the equations between the sections indicate which equation to use, depending on which way you are moving around the frequency tree.


RP
Problem Reference Population Section
Problem Reference Population
Eq. 1
Eq. 2
Eq. 3
Eq. 4
H1
H2
Hypothesis Section
Chance of Hypothesis Happening
Chance of Hypothesis Not Happening
Eq. 5
Eq. 6
Eq. 7
Eq. 8
Eq. 9
Eq. 10
Eq. 11
Eq. 12
E1
E2
E3
E4
Evidence/ Data Section
Chance of Evidence/ Data Supporting
Chance of Evidence/ Data Not Supporting
Chance of Evidence/ Data Supporting
Chance of Evidence/ Data Not Supporting

Equations:

The following is a list of the equations as referenced in the diagram above.

Equation 1:

H1 = RP – H2

Equation 2:

RP = H1/ Data given in problem for the value of H1 in decimal format or if H2 is known, (100 – percentage value of H2) in decimal format

Equation 3:

H2 = RP – H1

Equation 4:

RP = H2/ Data given in problem for the value of H2 in decimal format or if H1 is known, (100 – percentage value of H1) in decimal format

Equation 5:

E1 = H1 – E2

Equation 6:

H1 = E1/ Data given in problem for the value of E1 in decimal format or if E2 is known, (100 – percentage value of E2) in decimal format

Equation 7:

E2 = H1 – E1

Equation 8:

H1 = E2/ Data given in problem for the value of E2 in decimal format or if E1 is known, (100 – percentage value of E1) in decimal format

Equation 9:

E3 = H2 – E4

Equation 10:

H2 = E3/ Data given in problem for the value of E3 in decimal format or if E4 is known, (100 – percentage value of E4) in decimal format

Equation 11:

E4 = H2 – E3

Equation 12:

H2 = E4/ Data given in problem for the value of E4 in decimal format or if E3 is known, (100 – percentage value of E3) in decimal format

Annex: Bayesian Equation And Verification Of Example Case Answer

$$P(H|D) = \frac{P(H)*P(D|H)}{(P(H)*P(D|H))+(P(NH)*P(D|NH))}$$

Where:

P(H|D) is the probability that the hypothesis is true given the data.

P(H) is the probability of the hypothesis being true.

P(D|H) is the probability of the data given that the hypothesis is true (hit rate).

P(NH) is the probability that the hypothesis is not true.

P(D|NH) is the probability of the data given that the hypothesis is not true (false positive).

$$P(H|D) = \frac{0.02 * (0.95)}{(0.02 * 0.95) + (0.98 * 0.03)}$$

$$P(H|D) = \frac{0.019}{(0.019) + (0.0294)}$$

$$P(H|D) = \frac{0.019}{0.0484}$$

$$P(H|D) = 0.39$$

APPENDIX H

<u>Frequency Tree Method Practice Question and Answer</u>

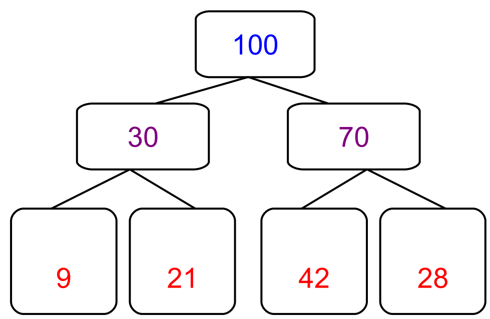
Practice Question

According to a recent analytical report, you know there is a 70 percent chance that Israel will break its truce with the Palestinians. While sifting through your daily traffic, you see a report that states Israel does not plan to break its truce with Palestine. Because of your familiarity with the source, you know there is a 30 percent chance that it is correct. In addition, you assess that there is a 60 percent chance it is being used as deliberate deception since most other sources indicate Israel will break its truce with Palestine. What are the approximate odds that Israel will not break its truce with the Palestinians?

Practice Question Answer

According to a recent analytical report, you know there is a 70 percent chance that Israel will break its truce with the Palestinians. While sifting through your daily traffic, you see a report that states Israel does not plan to break its truce with Palestine. Because of your familiarity with the source, you know there is a 30 percent chance that it is correct. In addition, you assess that there is a 60 percent chance it is being used as deliberate deception since most other sources indicate Israel will break its truce with Palestine. What are the approximate odds that Israel will not break its truce with the Palestinians?

Approximately 18%



9/51 = .176 = Approximately 18%

APPENDIX I

Test Questions and Answers

Test Questions

Please circle your answer and show as much work as possible. All information contained in the following questions is ***fictitious*** and is for training purposes only.

1. According to a recent analytical report, you know there is a 4 percent chance that Iran is using its civilian nuclear program to build nuclear weapons. While sifting through your daily traffic, you come across a SIGINT report indicating Iran is creating nuclear weapons. From previous experience, you know that a SIGINT report from this source is 60 percent accurate. However, being the analyst you are, you assess there is an 8 percent chance that it is deliberate deception. What are the approximate odds that Iran is actually using its civilian nuclear program to build nuclear weapons?

2. While sifting through your daily traffic, you come across a news article stating that politicians and renegade army officers loyal to the exiled Prime Minister staged the New Year's Eve bombings in Thailand. From experience you know that this news source is wrong about 10 percent of the time. In addition, you assess that there is a 5 percent chance the article is deliberate deception. However, this article has piqued your interest, so you take a look at analytical reports coming from the Asian regional office to see what they have to say. According to these analytical reports, you know there is an 80 percent that the politicians and renegade officers staged the bombings. What are the approximate odds that the politicians and renegade officers loyal to the exiled Prime Minister actually staged the New Year's Eve bombings in Thailand?

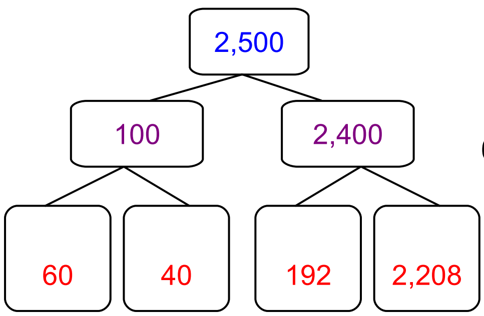
3. While reading your daily traffic for Latin America, you come across a news article that states Fidel Castro does not have cancer. However, recent analysis from the Cuban analysts suggests there is a 40 percent chance that Fidel Castro does have cancer. You know that this news article is likely being used by the Cuban government as deception and assess that there is a 60 percent chance this is deliberate deception. In addition, you also know that this news source is correctly only about 50 percent of the time. What are the approximate odds that Fidel Castro does not have cancer?

4. Recently, the Somali troops captured the last major stronghold of a militant Islamic movement in Kismayo. This caused hundreds of Islamic fighters to leave the town. However, in a recent report, you read that the Islamic fighters are re-grouping and there is a 10 percent chance that they will be able to re-take Kismayo. Being the analyst you are, you question the report and assess that there is a 75 percent chance that it is not deliberate deception, but know that this report is only right about 15 percent of the time. In addition, according to your fellow analysts and all published documents within your agency, there is actually only a 5 percent chance that the Islamic fighters are regrouping and planning to re-take Kismayo. What are the approximate odds that the Islamic fighters are regrouping and will re-take Kismayo?

Test Question Answers

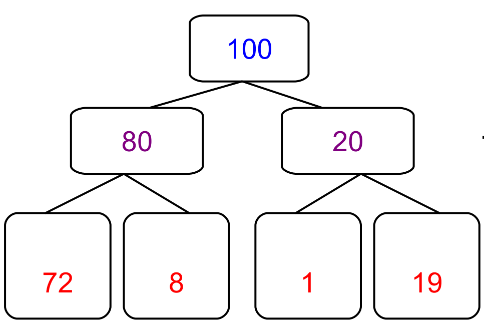
All information contained in the following questions is ***fictitious*** and is for training purposes only.

1. According to a recent analytical report, you know there is a 4 percent chance that Iran is using its civilian nuclear program to build nuclear weapons. While sifting through your daily traffic, you come across a SIGINT report indicating Iran is creating nuclear weapons. From previous experience, you know that a SIGINT report from this source is 60 percent accurate. However, being the analyst you are, you assess there is an 8 percent chance that it is deliberate deception. What are the approximate odds that Iran is actually using its civilian nuclear program to build nuclear weapons? **Approximately 24%**



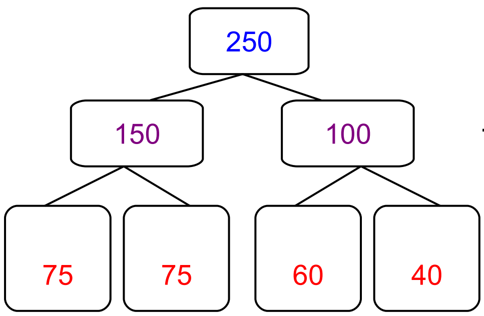
60/252 = .238 = Approximately 24%

2. While sifting through your daily traffic, you come across a news article stating that politicians and renegade army officers loyal to the exiled Prime Minister staged the New Year's Eve bombings in Thailand. From experience you know that this news source is wrong about 10 percent of the time. In addition, you assess that there is a 5 percent chance the article is deliberate deception. However, this article has piqued your interest, so you take a look at analytical reports coming from the Asian regional office to see what they have to say. According to these analytical reports, you know there is an 80 percent that the politicians and renegade officers staged the bombings. What are the approximate odds that the politicians and renegade officers loyal to the exiled Prime Minister actually staged the New Year's Eve bombings in Thailand? **Approximately 99%**



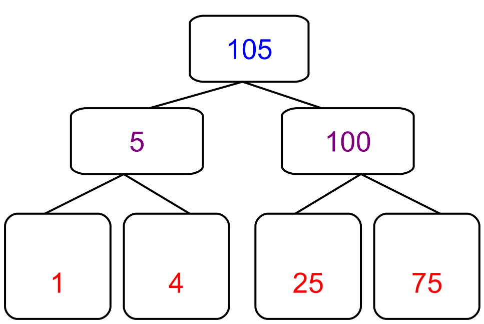
72/73 = .986 = Approximately 99%

3. While reading your daily traffic for Latin America, you come across a news article that states Fidel Castro does not have cancer. However, recent analysis from the Cuban analysts suggests there is a 40 percent chance that Fidel Castro does have cancer. You know that this news article is likely being used by the Cuban government as deception and assess that there is a 60 percent chance this is deliberate deception. In addition, you also know that this news source is correctly only about 50 percent of the time. What are the approximate odds that Fidel Castro does not have cancer? **Approximately 56%**



75/135 = .555 = Approximately 56%

4. Recently, the Somali troops captured the last major stronghold of a militant Islamic movement in Kismayo. This caused hundreds of Islamic fighters to leave the town. However, in a recent report, you read that the Islamic fighters are re-grouping and there is a 10 percent chance that they will be able to re-take Kismayo. Being the analyst you are, you question the report and assess that there is a 75 percent chance that it is not deliberate deception, but know that this report is only right about 15 percent of the time. In addition, according to your fellow analysts and all published documents within your agency, there is actually only a 5 percent chance that the Islamic fighters are regrouping and planning to re-take Kismayo. What are the approximate odds that the Islamic fighters are regrouping and will re-take Kismayo? **Approximately 4%**



1/26 = .038 = Approximately 4%

APPENDIX J

Post-Test Questionnaire

Frequency Tree Method and Bayesian Reasoning Follow-Up Questionnaire

Thanks for your participation! Please take a few moments to answer the following questions. Your feedback is greatly appreciated.

1. Please rate your understanding of Bayesian reasoning prior to this study, with 1 being extremely low and 5 being extremely high.

   1 2 3 4 5

2. Please rate your understanding of Bayesian reasoning following this study, with 1 being extremely low and 5 being extremely high.

   1 2 3 4 5

3. Please rate how easy the frequency tree method procedure was to follow, with 1 being not easy and 5 being extremely easy.

   1 2 3 4 5

4. Please rate how strong of math background you have, with 1 being not strong and 5 being very strong.

   1 2 3 4 5

5. Please rate how useful you feel a math background is for performing the frequency tree method, with 1 being not useful and 5 being very useful.

   1 2 3 4 5

6. Please rate how useful the training session was in explaining the frequency tree method and answering your questions, with 1 being not useful at all and 5 being extremely useful.

   1 2 3 4 5

7. Please rate how useful you think this method will be to you as an intelligence analyst, with 1 being not useful and 5 being extremely useful.

   1 2 3 4 5

8. Do you plan to use this method or Bayesian reasoning in general as an analytic tool?

    Please circle one yes or no for each.

    This Method: Yes No

    Bayesian Reasoning: Yes No

9. Please provide any additional comments you may have regarding this method, Bayesian reasoning, and the training session.

APPENDIX K

Participant Debriefing Statement

Bayesian Reasoning
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 test the viability of my frequency tree method for Bayesian reasoning. Previous research indicates that Bayesian reasoning is easier for people to perform when information is presented and calculated using natural frequencies. Natural frequencies, simply defined, put percentages in terms that are easily understandable to those who do not have strong statistical backgrounds. My frequency tree method converts statistical information into natural frequencies to solve intelligence problems using Bayesian reasoning.

Bayesian reasoning has long been discussed as a possible analytical tool that would be useful for the intelligence community. I plan to use the results from this study to support Bayesian reasoning and introduce a new method that intelligence analysts will be able to use to solve Bayesian problems.

If you have any further question about Bayesian reasoning or this research you can contact me at XXX.

APPENDIX L

Email Sent to All Students in the Intelligence Studies Program at Mercyhurst College for the First Natural Frequency Tree Method Experiment

To All Intelligence Studies Students,

Jennifer Lee, a second year graduate student in the intelligence program, has developed a Bayesian Reasoning method for her thesis and is looking for participants to take the time to learn her method and provide feedback regarding this analytical technique. This is the **SECOND TIME** she is running her experiment and is looking for students who **did not participate** the first time.

Bayesian Reasoning is a great technique to understand. It will make you a better analyst and give you a competitive edge when it comes to talking to employers. Participation involves learning her method to perform Bayesian Reasoning, applying the method to a few problems, and filling out a questionnaire.

- You **do not need** to have a strong mathematical background to perform this analytical technique.

- The total time commitment required for your participation is **1 ½ hours**.

- Professors Mills (Advanced Class Only), Breckenridge, Grabelski, Wheaton, Mulligan, Welch and Wozneak have agreed to offer **extra credit** for your participation.

This is your chance to gain an understanding of Bayesian Reasoning. If you are interested in participating, **please let Jen know via email at XXX** by 2:00pm tomorrow (April 11), for the session that will be held on Tuesday, April 11 at 5:30 PM in Zurn 213.

Please **DO NOT** respond to the RIAP group or to me directly. Send your notes to Jen.

Prof. Wheaton

APPENDIX M

Revised Frequency Tree Method Procedures with Changes Highlighted in Yellow

Frequency Tree Method Instructions

By: Jennifer Lee

Tools needed:

Paper
Pencil or Pen
Calculator

Training Case:

This example is comprised solely of ***fictitious*** information and is for training purposes only.

According to a recent analytical report, you know there is a 2 percent chance Japan is creating nuclear weapons. While sifting through your daily traffic, you come across a SIGINT report indicating Japan is stockpiling plutonium for creating nuclear weapons. From previous experience, you know that a SIGINT report from this source is 95 percent accurate. However, being the analyst you are, you assess there is a 3 percent chance that it is deliberate deception.

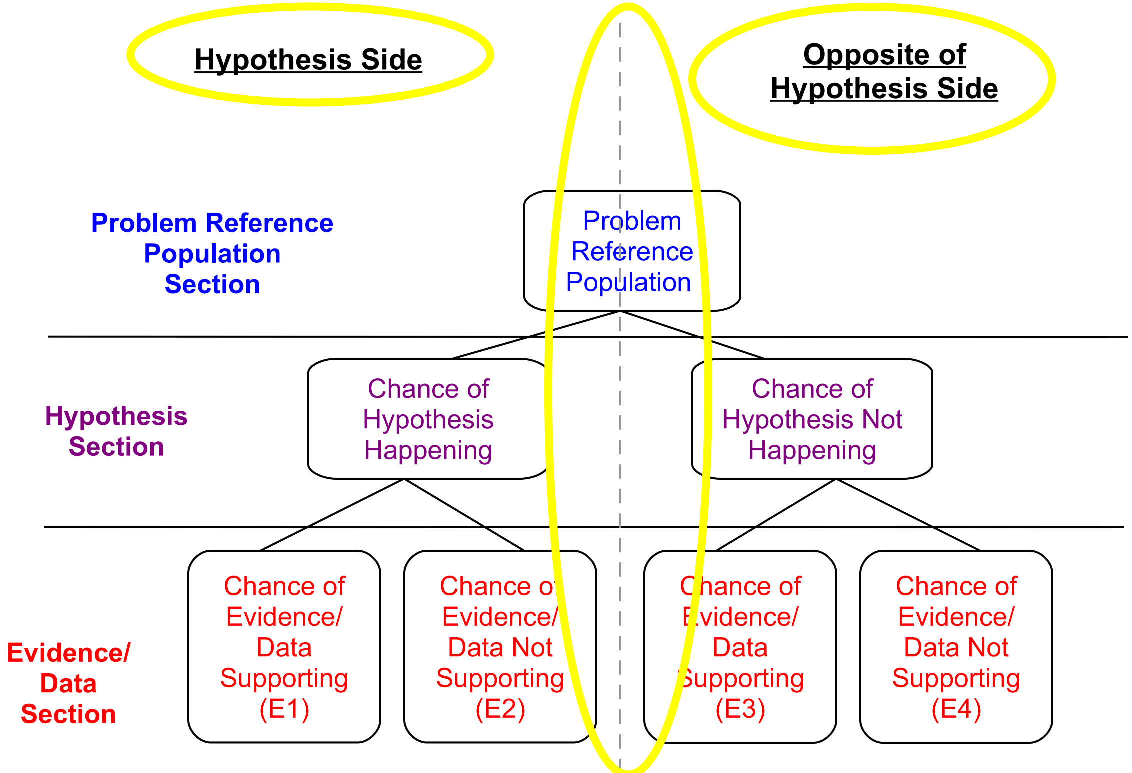
What are the approximate odds that Japan is actually stockpiling plutonium for the creation of nuclear weapons?

Frequency Tree Steps:

4. Draw Frequency Tree
5. Create Natural Frequencies And Fill In Data
6. Calculate Answer

Step 1: Draw Frequency Tree

The frequency tree is comprised of three sections (see next page):



The frequency tree can be broken down into two sides: a hypothesis side and an opposite of hypothesis side. ***It is very important to keep this in mind when deciding where to place the numbers in the tree.***

The Problem Reference Population section will most likely not be given in the problem data. This section will be calculated from the information in the Hypothesis section.

The Hypothesis section includes the information from the problem that describes the chances of the hypothesis (main event or action) happening or not happening (opposite of main action or event). This is usually what your new evidence suggests.

The Evidence/Data section includes the information that acts as evidence of the hypothesis happening or not happening. The evidence can either support the chances of the hypothesis happening or it can support the chances of the opposite of your hypothesis happening.

For each problem, draw an empty frequency tree. Step 2 will discuss how to fill in the frequency tree using the data from the problem.

Check List Before Moving To Next Step:

Draw empty frequency tree.

Step 2: Create Natural Frequencies And Fill In Data

Natural frequencies, simply defined, put percentages in terms that are easily understandable to those who do not have strong statistical backgrounds. What natural frequencies provide are the reference populations to the percentages. For example, 10 percent can easily be put into a natural frequency by saying "10 out of 100". "10 out of 100" is equal to 10 percent and more easily understood. In addition, "10 out of 100" provides information about the reference population, which is 100 in this example.

Step 2A:

Before creating the natural frequencies, determine where the ***given information*** you currently have from the problem fits into the frequency tree. It is also very important to identify at this time what the hypothesis is, or the question you are trying to answer. This is the ***most important*** step, as it identifies where in the frequency tree each piece of goes. The best thing to do is to write out next to your frequency tree what each piece of information represents, as shown below. Keep in mind that each section is the reference population for the section below it. Below are the pieces of information from the problem above. They have been listed and labeled as to where they fit into the frequency tree:

Hypothesis: Japan is stockpiling plutonium for the creation of nuclear weapons.

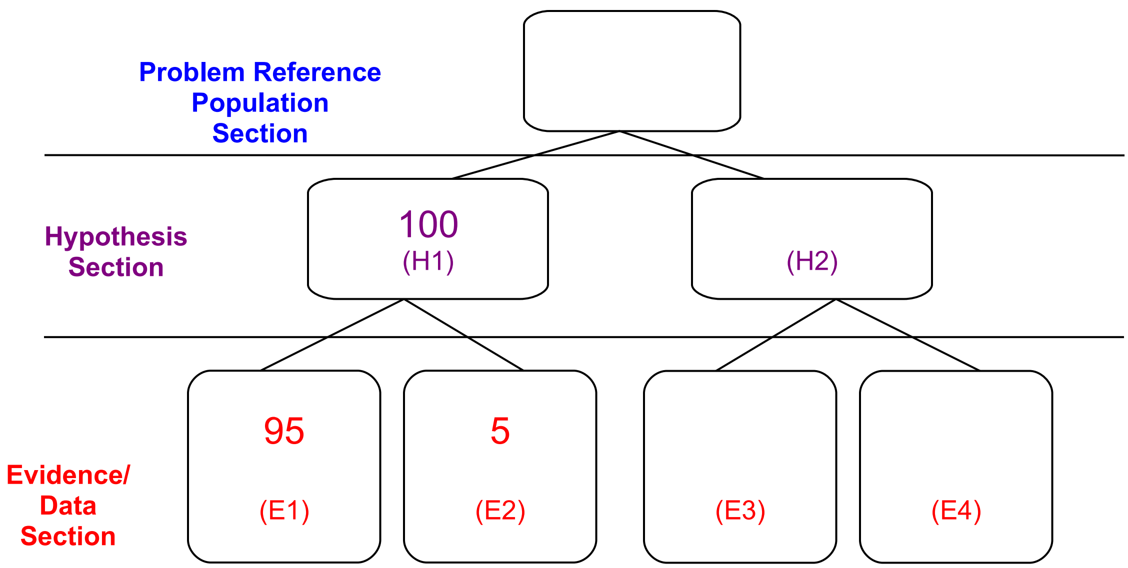
Hypothesis Section (H1): There is a 2 percent chance that Japan is stockpiling plutonium for the creation of nuclear weapons.

Evidence/Data Section (E1): A SIGINT report indicates Japan is creating nuclear weapons. From previous experience, you know that a SIGINT report from this source is 95 percent accurate.

Evidence/Data Section (E3): You assess there is a 3 percent chance that it is deliberate deception.

To put statistical information into natural frequencies easily, skim the problem for the largest given percentage. In the example above, that would be 95. To keep 95 a whole number, its reference population needs to be 100. The reference population for the first number will always be 100. The 95 goes into the box E1 in the Evidence/Data section (as described above), and since the level above it is its reference population, the 100 goes into the Hypothesis section's H1 box.

Keep in mind that the largest number ***will not*** necessarily always go in box E1 in the Evidence/Data section. Depending on the problem, the largest number could go in any box in the Hypothesis section or Evidence/Data section. Another key point to remember is that the boxes in the section below the reference population always need to add up to the reference population value. In this problem, since boxes E1 and E2 underneath box H1 need to add up to the reference population, which in this problem is 100, the value that goes into box E2 in the Evidence/Data section is 5.



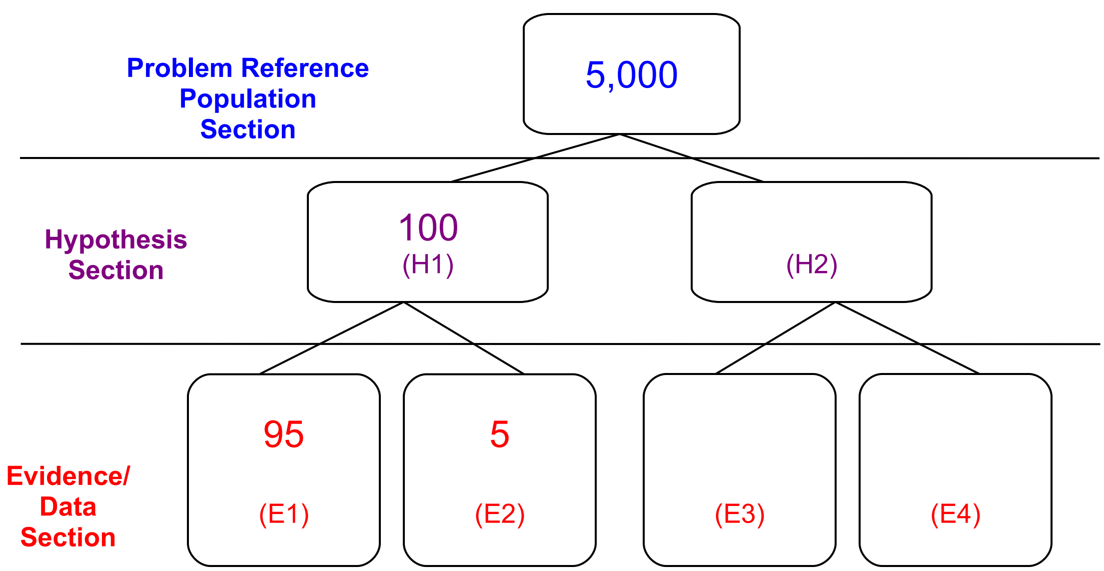
Check List Before Moving To Next Step:

- Write down each piece of information from the problem and where each piece fits into frequency tree.
- Put the largest given number into the correct box in the frequency tree and made the value (reference population) in the box above it 100 according to the information given in the problem.

Step 2B:

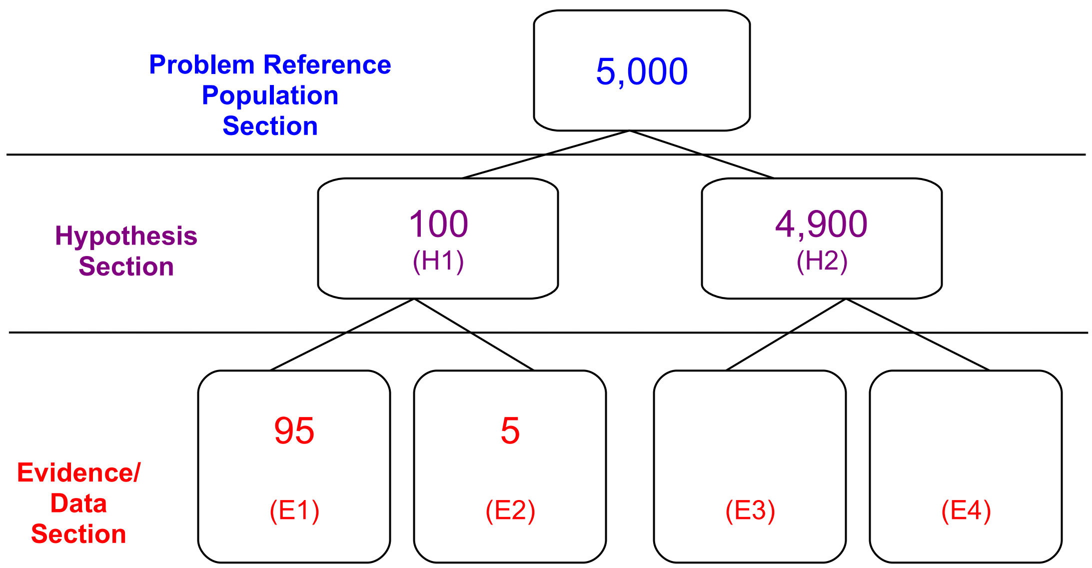
Once the first piece of information is filled-in, the next step is to determine the reference population for the box above the filled-in boxes. In the above example, this would be determining the reference population for the 100 in the Hypothesis section. From the information in the problem, we know that the 100 is equal to 2 percent of the Problem Reference Population. To determine the Problem Reference Population, divide the Hypothesis section reference population in box H1, which is the 100, by 2 percent, or 0.02. Doing this simple calculation provides the Problem Reference Population, which is 5,000.

It is important to remember that it does not matter what the reference population turns out to be as long as the ratios between the reference population and the boxes below it match. It is ***very important*** to keep the ratios correct. In this example, the reference population is 5,000, but in another problem, depending on the ratios, the reference population could be 2,000. Additionally, all of the reference populations should be whole numbers, even if this means having to round up or down. Since you are most likely only looking for an approximate answer to the question, rounding should not be an issue.



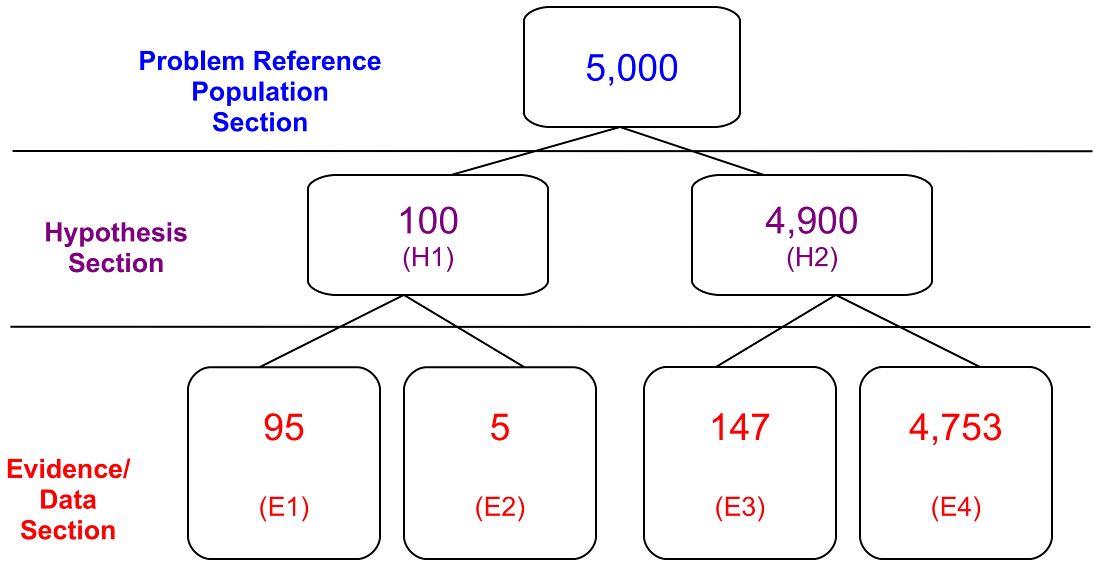
Step 2C:

The next step is to fill in box H2 in the Hypothesis section. Since we know that the two Hypothesis section boxes need to add up to 5,000, subtract the 100 in left box H1 from the 5,000 in the Problem Reference Population section. This gives you 4,900.



Step 2D:

The final step is to calculate the values for boxes E3 and E4 in the Evidence/Data section. To do this, we need to use the last piece of evidence, which states that 3 percent of the time, there is deliberate deception (also known as the chance of the evidence supporting the main event or action happening even though the main event or action will not be happening). This number goes in box E3 in the Evidence/Data section. To determine the number, multiply 4,900 by 3 percent, or 0.03. This calculation results in 147. Once again, we know that boxes E3 and E4 in the Evidence/Data section need to add up to the reference population in Hypothesis section box H2. Therefore, we need to subtract 147 from 4,900, which results in 4,753.

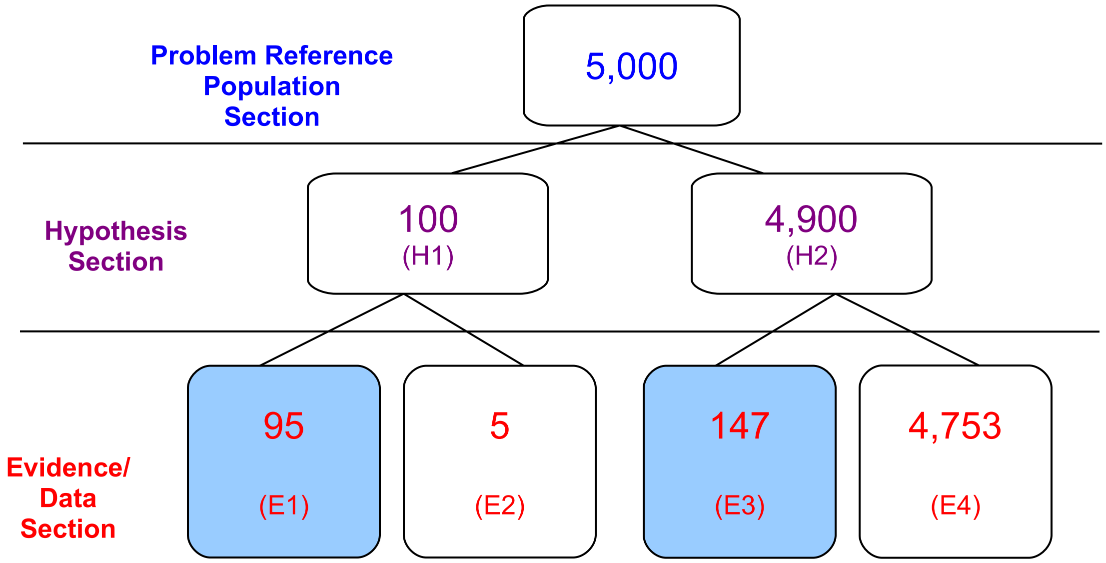


Check List Before Moving To Next Step:

- Make sure the ratios between the reference population(s) and the boxes below the reference population(s) match the information given in the problem.
- Make sure the values in the boxes below the reference population(s) add up correctly.

Step 3: Calculate Answer

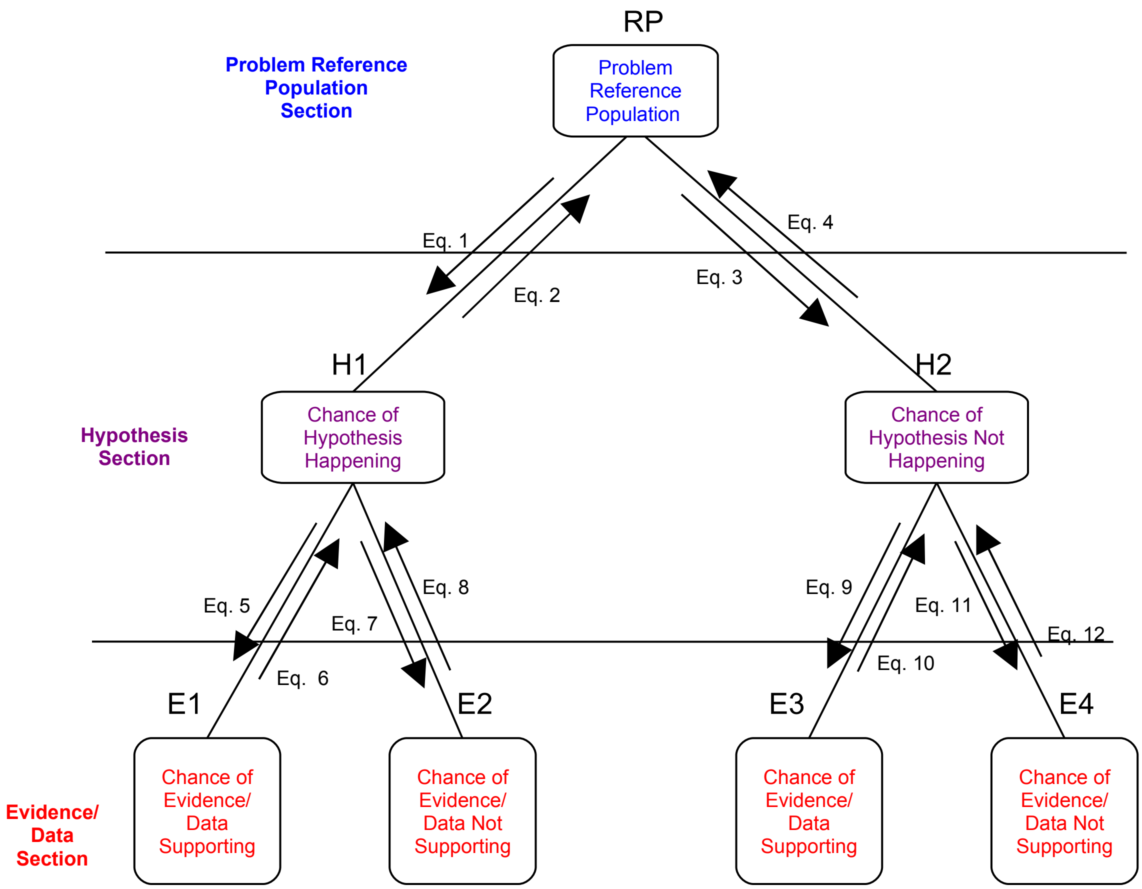
The key to calculating the answer is boxes E1 and E3 (shaded blue) in Evidence/Data section of the frequency tree.



To determine the answer to the question, "What are the approximate odds that Japan is creating nuclear weapons given this new piece of evidence?" all you need to do is divide box E1 (95) by the sum of boxes E1 and E3 (95 plus 147). Therefore, the answer to the question is 95 divided by 242 (the sum of 95 plus 147), which is .39 or 39 percent. To verify this answer, one can simply input the data into the real Bayesian formula (see appendix).

Mathematical Calculations To Move Around The Frequency Tree

As mentioned above, the largest input number (the number you initially start with) will not always be the box E1 in the Evidence/Data section. The frequency tree method allows you to input initial data into any of the Hypothesis section or Evidence/Data section boxes. The key to using the frequency tree method is correctly determining which section the given information falls into (Step 2 above). Once that has been determined, you can easily calculate the values of the boxes using the calculations below. As described in the steps above, always start with the largest number. Also, do not forget that each section's box values need to add up to the reference population in the section above it. In the diagram below, each box has a reference (that corresponds to the box references used above) above it since the calculations will refer to specific boxes using these references. The arrows next to the equations between the sections indicate which equation to use, depending on which way you are moving around the frequency tree.



RP
Problem Reference Population Section
Problem Reference Population
Eq. 1
Eq. 2
Eq. 3
Eq. 4
H1
H2
Hypothesis Section
Chance of Hypothesis Happening
Chance of Hypothesis Not Happening
Eq. 5
Eq. 6
Eq. 7
Eq. 8
Eq. 9
Eq. 10
Eq. 11
Eq. 12
E1
E2
E3
E4
Evidence/ Data Section
Chance of Evidence/ Data Supporting
Chance of Evidence/ Data Not Supporting
Chance of Evidence/ Data Supporting
Chance of Evidence/ Data Not Supporting

Equations:

The following is a list of the equations as referenced in the diagram above.

Equation 1:

$H1 = RP – H2$

Equation 2:

RP = H1/ Data given in problem for the value of H1 in decimal format or if H2 is known, (100 – percentage value of H2) in decimal format

Equation 3:

$H2 = RP – H1$

Equation 4:

RP = H2/ Data given in problem for the value of H2 in decimal format or if H1 is known, (100 – percentage value of H1) in decimal format

Equation 5:

$E1 = H1 – E2$

Equation 6:

H1 = E1/ Data given in problem for the value of E1 in decimal format or if E2 is known, (100 – percentage value of E2) in decimal format

Equation 7:

E2 = H1 – E1

Equation 8:

H1 = E2/ Data given in problem for the value of E2 in decimal format or if E1 is known, (100 – percentage value of E1) in decimal format

Equation 9:

E3 = H2 – E4

Equation 10:

H2 = E3/ Data given in problem for the value of E3 in decimal format or if E4 is known, (100 – percentage value of E4) in decimal format

Equation 11:

E4 = H2 – E3

Equation 12:

H2 = E4/ Data given in problem for the value of E4 in decimal format or if E3 is known, (100 – percentage value of E3) in decimal format

Annex: Bayesian Equation And Verification Of Example Case Answer

$$P(H|D) = \frac{P(H)*P(D|H)}{(P(H)*P(D|H))+(P(NH)*P(D|NH))}$$

Where:

P(H|D) is the probability that the hypothesis is true given the data.

P(H) is the probability of the hypothesis being true.

P(D|H) is the probability of the data given that the hypothesis is true (hit rate).

P(NH) is the probability that the hypothesis is not true.

P(D|NH) is the probability of the data given that the hypothesis is not true (false positive).

$$P(H|D) = \frac{0.02 * (0.95)}{(0.02 * 0.95) + (0.98 * 0.03)}$$

$$P(H|D) = \frac{0.019}{(0.019) + (0.0294)}$$

$$P(H|D) = \frac{0.019}{0.0484}$$

$$P(H|D) = 0.39$$

APPENDIX N

Revised Frequency Tree Method Practice Question 1 and Answer

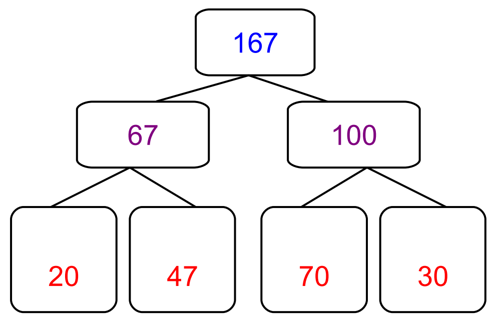
Practice Question 1

According to a recent analytical report, you know there is a 60 percent chance that Israel will break its truce with the Palestinians. While sifting through your daily traffic, you see a report that states Israel does not plan to break its truce with Palestine. Because of your familiarity with the source, you know there is a 30 percent chance that it is correct. In addition, you assess that there is a 70 percent chance it is being used as deliberate deception since most other sources indicate Israel will break its truce with Palestine. What are the approximate odds that Israel will not break its truce with the Palestinians?

Practice Question 1 Answer

According to a recent analytical report, you know there is a 60 percent chance that Israel will break its truce with the Palestinians. While sifting through your daily traffic, you see a report that states Israel does not plan to break its truce with Palestine. Because of your familiarity with the source, you know there is a 30 percent chance that it is correct. In addition, you assess that there is a 70 percent chance it is being used as deliberate deception since most other sources indicate Israel will break its truce with Palestine. What are the approximate odds that Israel will not break its truce with the Palestinians?

Approximately 22%



20/90 = .222 = Approximately 22%

APPENDIX O

Revised Frequency Tree Method Practice Question 2 and Answer

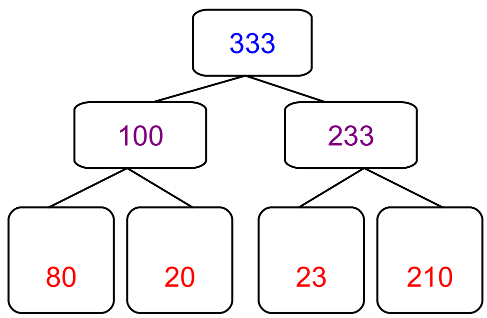
Practice Question 2

While sifting through your daily traffic, you come across an article that states Egypt is now becoming a center for third-party and contracted torture. You know that the Egyptian government denies that this is happening. After reading a recent analytical report put out by another area in your organization, you conclude that there is a 30 percent chance Egypt is actually performing torture. You assess that the article you just read from your daily traffic only has a 10 percent chance of being deliberate deception. You also know that the source for the article is very reliable and is correct about 80 percent of the time. What are the approximate odds that Egypt is actually becoming a center for third-party and contract torture?

Practice Question 2 Answer

While sifting through your daily traffic, you come across an article that states Egypt is now becoming a center for third-party and contracted torture. You know that the Egyptian government denies that this is happening. After reading a recent analytical report put out by another area in your organization, you conclude that there is a 30 percent chance Egypt is actually performing torture. You assess that the article you just read from your daily traffic only has a 10 percent chance of being deliberate deception. You also know that the source for the article is very reliable and is correct about 80 percent of the time. What are the approximate odds that Egypt is actually becoming a center for third-party and contract torture?

Approximately 78%



80/103 = .776 = Approximately 78%

APPENDIX P

Natural Frequency Verification Experiment Complete Statistical Findings and Raw Data

Complete Statistical Findings

Question 1: Is there a significant difference between the Traditional Statistics group and the Natural Frequencies group?

Hypothesis Statements:

Null Hypothesis: There is no difference between using the traditional statistics language approach and using the natural frequency approach.

Alternate Hypothesis: There is a significant difference between using the traditional statistics language approach and using the natural frequency approach. (Claim)

Assumption Checking:

For both groups, the sample size is large (30 or more).

Independent samples are present, as a different approach was used for each sample/data set.

Decision is to use a Z – distribution two-hypothesis test.

Analysis:

Table A.1: Group Statistics For The Natural Frequency Verification Experiment Traditional Group And Natural Frequency Group

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Frequency per group	List of Answers from Problem using Percentages	34	38.3382	37.78823	6.48063
	List of Answers from Problem using Natural Frequencies	33	13.4606	25.09408	4.36832

Table A.2: Complete Independent Samples Test Results For The Z-Distribution Two-Hypothesis Test (Natural Frequency Verification Experiment)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Frequency per group	Equal variances assumed	17.876	.000	3.164	65	.002	24.87763	7.86150	9.17713	40.57812
	Equal variances not assumed			**3.183**	57.548	**.002**	24.87763	7.81542	9.23076	40.52450

Z – test value = 3.183
P-value = 0.002
P-value = 0.002 is smaller than the level of significance of 0.05, thus reject the null hypothesis.

<u>Conclusion</u>:

At the 5% level, there is a significant difference between using the traditional statistics language approach and using the natural frequency approach.

Question 2: Is there a significant difference between Hoffrage and Gigerenzer's doctor data set and the Intelligence Studies data set for the Traditional Statistics group?

Hypothesis Statements:

Null hypothesis: There is no difference between Hoffrage and Gigerenzer's doctor data set and the Intelligence Studies data set for the Traditional Statistics group.

Alternative hypothesis: There is a significant difference between Hoffrage and Gigerenzer's doctor data set and the Intelligence Studies data set for the Traditional Statistics group. (Claim)

Assumption Checking:

Sample size for Hoffrage and Gigerenzer's doctor data is small (less than 30).

Samples are independent, as the data is collected from two different professions.

Normality assumption check:

Null: The variable is normally distributed.

Alternative: The variable is not normally distributed.

Table A.3: Normality Test Results For Hoffrage And Gigerenzer's Doctor Sample

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
List of Answers from Problem using Percentages	Doctor data with Traditional Statistical Method	.241	24	**.001**	.769	24	.000
	Intel Student data with Traditional Statistical Method	.219	34	**.000**	.823	34	.000

a. Lilliefors Significance Correction

From the Kolmogorov-Smirnov column in the above output table, for both the samples P-values are less than 0.05; thus, reject the null hypothesis. The normality assumption is not satisfied; however, normality is robust and we should is to check the plots.

Figure A.1: Normality Plot 1 For Hoffrage And Gigerenzer's Doctor Sample

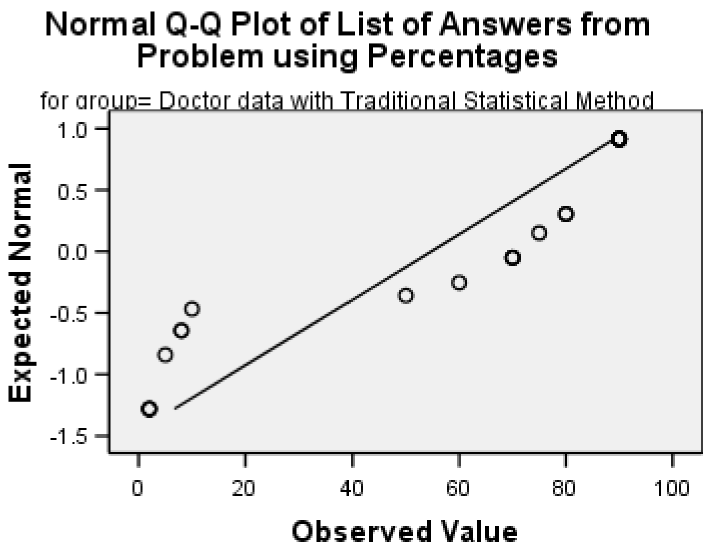
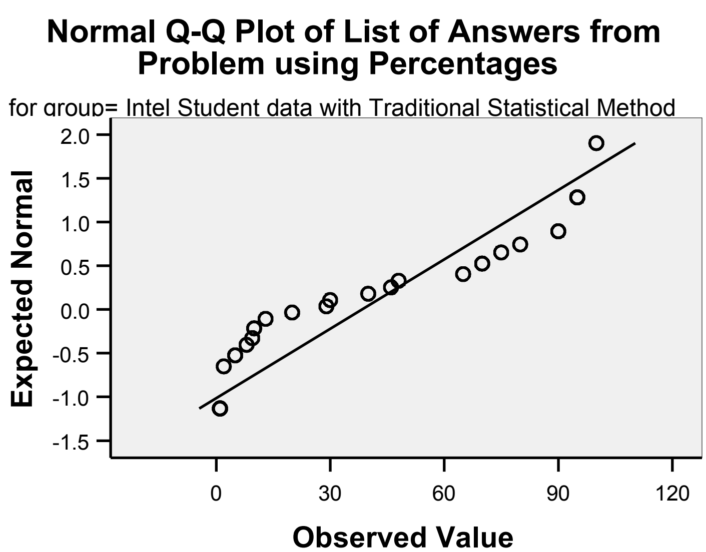


Figure A.2: Normality Plot 2 For Hoffrage And Gigerenzer's Doctor Sample



Both Normal Q-Q plots show that most of the points are close to the diagonal line. Thus, the assumption of normality is satisfied for the samples.

Note that for the doctor data, a few points (lower left) look as if they are away from the line, but if you look at the vertical axis scale, it really does not matter.

Decision is to use a t-distribution independent two-hypothesis test.

Analysis:

Table A.4: Group Statistics For Hoffrage And Gigerenzer's Doctor Data And The Natural Frequency Verification Experiment's Traditional Statistics Group

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
List of Answers from Problem using Percentages	Doctor data with Traditional Statistical Method	24	54.7500	37.50507	7.65569
	Intel Student data with Traditional Statistical Method	34	38.3382	37.78823	6.48063

Table A.5: Complete Doctors vs. Analysts Independent Samples Test Results For The t-Distribution Independent Two-Hypothesis Test

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
List of Answers from Problem using Percentages	Equal variances assumed	.010	**.920**	**1.634**	56	**.108**	16.41176	10.04361	-3.70800	36.53153
	Equal variances not assumed			1.636	49.911	.108	16.41176	10.03036	-3.73570	36.55923

In order to decide the t-test value, we need to see if the equal variance assumption is satisfied.

Null: The variances are equal.

Alternative: The variances are not equal.

According to Levene's test in the above output table, the P-value is 0.920. It is larger than the level of significance of 0.05. Thus, we fail to reject the null hypothesis.

Therefore, the variances are equal at 5% level.

t-test value = 1. 634
P-value = 0.108

The P-value of 0.108 is larger than level of significance of 0.05, thus we fail to reject the null hypothesis.

Conclusion:

There is no difference between Hoffrage and Gigerenzer's doctor data set and the Intelligence Studies data set for the Traditional Statistics group at the 5% level.

Raw Data

Table A.6: List of Answers for the Traditional Statistical Language Group and the Natural Frequencies Group

Traditional Statistical Language Group N = 34	Natural Frequencies Group N= 33
1	0.1
1	0.1
1	1
1	1
1	1
1	2
1	2
1	2
2	2
5	2
5	2
8	2
9.5	2
10	2
10	2
13	2
20	2
29	2
30	5
40	5
46	5
48	5
65	5
70	5
70	10
75	10
80	20
90	20
90	25
95	45
95	65
95	95
95	95
100	

APPENDIX Q

Natural Frequency Tree Method Experiment Complete Statistical Findings and Raw Data

Complete Statistical Findings

Question: Is there a significant difference between the results of the Natural Frequency group (79% baseline) from the natural frequency verification experiment and the results obtained from Test Question 1 in the natural frequency tree method experiment?

Hypothesis Statements:

Null: There is no difference between the Natural Frequency group results from the natural frequency verification experiment and Test Question 1 results from the natural frequency tree method experiment.

Alternative: There is a significant difference between the Natural Frequency group results from the natural frequency verification experiment and Test Question 1 results from the natural frequency tree method experiment. (Claim)

Assumption Checking:

The assumption of normality is satisfied as the sample size is large, i.e. the sample size is at least 30.

Decision is to perform a One-Sample Z-distribution Hypothesis test.

Analysis:

Table A.7: Sample Statistics For The Natural Frequency Verification Sample Population

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Question 1	92	.7609	.42889	.04471

Table A.8: Original Natural Frequency Tree Method Experiment One-Sample Test Results For The Z-Distribution One-Sample Hypothesis Test

One-Sample Test

	Test Value = .79					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Question 1	**-.651**	91	**.516**	-.02913	-.1180	.0597

SPSS does not have a Z-test, but if the degrees of freedom are larger than 30, a t-distribution approaches a z-distribution; thus it is OK to use a one-sample t-test in SPSS.

Z-test value = -0.651
P-value = 0.516
P-value = 0.516 is greater than the level of significance 0.05, thus we fail to reject the null hypothesis.

Conclusion:

At the 5% level, the results from Test Question 1 are not significantly different than the results obtained from Natural Frequency group in the natural frequency verification experiment.

Raw Data

Table A.9: Correct Number Of Test Question Answers[150] Within 10 Points Of Actual Answer N = 92				
Key: 1 = Answered Question Correctly Blank = Answered Question Incorrectly				
	Test Question 1	Test Question 2	Test Question 3	Test Question 4
Participant				
1				1
2		1		
3	1	1		1
4				
5	1	1		1
6	1		1	
7			1	
8	1	1	1	1
9	1	1	1	1
10	1	1		
11	1	1		
12	1	1		1
13	1	1	1	
14	1			1
15	1	1	1	
16	1	1		
17	1	1	1	1
18	1			
19	1		1	1
20	1	1	1	
21	1		1	1
22	1	1	1	1
23	1	1	1	
24	1	1		
25	1	1	1	1
26			1	1
27	1	1		
28	1	1	1	1
29	1	1	1	1
30	1	1	1	
	Test Question 1	Test Question 2	Test Question 3	Test Question 4

[150] See Appendix I for actual test questions.

Participant				
31	1	1	1	1
32	1	1	1	
33	1	1	1	1
34	1		1	1
35	1	1	1	
36	1	1	1	
37				1
38	1	1	1	
39	1	1	1	
40	1	1	1	1
41	1	1	1	
42	1		1	1
43	1	1		
44	1	1	1	1
45	1	1	1	
46	1			
47	1		1	1
48	1	1	1	
49	1	1		
50			1	1
51	1	1	1	1
52	1		1	
53	1		1	1
54	1			
55	1	1	1	1
56				
57	1		1	
58	1	1		
59	1	1	1	1
60	1	1		
61	1	1		
62	1	1		
63	1		1	
64	1	1	1	1
65	1	1		1
66			1	
67	1		1	
68		1	1	1
69		1		1
70	1	1	1	
71	1	1	1	1
72	1			
73				1
74	1	1	1	1
	Test Question 1	**Test Question 2**	**Test Question 3**	**Test Question 4**
Participant				
75				1
76				1

77	1	1	1	
78	1	1		1
79			1	
80			1	1
81	1	1	1	
82	1	1	1	1
83	1	1	1	1
84		1		1
85	1	1	1	1
86		1		1
87				
88	1	1	1	
89	1			
90		1		
91		1		
92		1		
Total Number Correct:	**70**	**61**	**55**	**44**
Percentage Correct:	**0.76087**	**0.663043**	**0.597826**	**0.478261**

APPENDIX R

Post-Test Questionnaire Complete Statistical Findings and Raw Data

Complete Statistical Findings

Question: Is there a significant difference between Questionnaire Question 1's average answer and Questionnaire Question 2's average answer?

Hypothesis Statements:

Null: There is no difference between the average answers of Questionnaire Question 1 and Questionnaire Question 2.

Alternative: There is a significant difference between the average answers of Questionnaire Question 1 and Questionnaire Question 2.

Assumption Checking:

Assumption of normality is satisfied as both the sample sizes are large, i.e. the sample size is at least 30.

Decision is to perform a t-test for dependent samples.

Analysis:

Table A.10: Original Natural Frequency Tree Method Experiment Post-Test Questionnaire Questions 1 And 2 Sample Statistics

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Questionnaire Question 1	1.1935	93	[illegible]	[illegible]
	Questionnaire Question 2	[illegible]	[illegible]	[illegible]	[illegible]

Table A.11: Complete Original Natural Frequency Tree Method Post-Test Questionnaire Paired Samples Test Results For The t-Test For Paired Samples

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Questionnaire Question 1 - Questionnaire Question 2	-1.95161	.91504	.09489	-2.14006	-1.76316	**-20.568**	92	**.000**

t-test value = -20.568
P-value = 0.000
P-value = 0.000 is smaller than the level of significance of 0.05, thus we reject the null hypothesis.

Conclusion:

At the 5% level, the results for Questionnaire Question 1 are significantly different than the results for Questionnaire Question 2.

Raw Data

Table A.12: Answers From Post-Test Questionnaire (Original Natural Frequency Tree Experiment									
Key: Questions 1-7: Scale Answers Ranging From 1 - 5 Questions 8A & 8B: 1 = Yes; Blank = No									
Question Number:	Q. 1	Q. 2	Q. 3	Q. 4	Q. 5	Q. 6	Q. 7	Q.8A	Q. 8B
Participant									
1	1	1	1	1	4	3	1		
2	1	3	4	4	2	4	5	1	1
3	1	3	2	3	2	1	5		1
4	5	3	4	4	3	4	3	1	
5	1	3	3	4	2	3	2		
6	1	3	2	4	2	3	2.5		
7	1	4	4	5	3	5	5	1	
8	1	4	4	2	2	2	5	1	1
9	1	3	4	2	2	3	3	1	1
10	1	3	4	2	4	4	4	1	1
11	1	2	1	2	3	1	1		
12	1	3	3	5	1	4	3	1	
13	1	3	4	5	5	3	5	1	
14	1	3	4	2	2	4	4	1	1
15	1	3	3	2	3	4	4		1
16	1	2	1	5	2	2	2		1
17	1	2	2	4	4	2	4	1	1
18	1	2	1	3	4	2	3	1	1
19	1	3	3	3	3	5	3	1	1
20	1	3	2	1	5	3	4		
21	1	3	3	2	2	4	3	1	1
22	1	4	4	3	3	5	4	1	1
23	2	4	4	4	3	5	4	1	1
24	4	4	4	4	4	4	4	1	1
25	1	3	3	2	3	5	4	1	1
26	1	3	3	2	2	4	4	1	1
Question Number:	Q. 1	Q. 2	Q. 3	Q. 4	Q. 5	Q. 6	Q. 7	Q.8A	Q. 8B

Participant									
27	1	4	5	4	2	5	4	1	1
28	1	3	3	3	2	4	4	1	
29	1	3	4	2	3	4	3	1	
30	1	4	3	4	3	4	3	1	
31	1	3	3	3	3	4	2		
32	1	3	3	3	2	4	4	1	
33	1	4	4	3	4	5	5	1	
34	1	4	4	1	3	3	4	1	1
35	1	3	4	4	2	3	4	1	
36	1	3	4	3	2	4	4	1	
37	1	3	2	1	4	2	2	1	
38	1	4	4	2	2	3	4	1	1
39	1	3	4	4	3	4	5	1	1
40	2	4	3	3	3	4	5	1	1
41	1	3	2	2	2	4	4	1	
42	1	3	4	3	2	4	4	1	1
43	1	5	5	5	3	5	4	1	1
44	1	3	3	3	4	4	3	1	
45	1	3	3	4	2	3	4		
46	1	3	4	4	3	5	5		1
47	1	2	2	3	3	3	2		
48	4	4	4	4	3	4	5	1	1
49	1	4	4	5	4	5	5	1	1
50	1	4	3	3	2	4	4	1	1
51	1	4	4	4	4	4	5	1	
52	1	4	3	1	3	4	5	1	
53	1	3	4	5	4	5	5	1	1
54	1	4	4	2	2	5	5	1	1
55	3	4	3	4	2	4	3	1	1
56	1	4	4	5	4	4	5	1	1
57	1	4	3	4	3	4	4	1	
58	1	3	4	2	3	5	4	1	
59	1	3	3	4	1	3	3		
Question Number:	**Q. 1**	**Q. 2**	**Q. 3**	**Q. 4**	**Q. 5**	**Q. 6**	**Q. 7**	**Q.8A**	**Q. 8B**
Participant									
60	1	3	3	3	3	4	3		

61	1	2	2	4	3	2	2		1
62	1	3	3	4	3	5	4	1	
63	1	3	3	4	3	3	2	1	
64	2	4	3	5	3	3	4	1	1
65	1	4	3	3	4	4	4	1	1
66	1	3	4	5	4	5	5		1
67	1	3.5	3	4	4	4	2.5	1	
68	1	3	3	1	2	3	3		
69	1	3	2	2	2	3	4	1	1
70	1	2	3	3	3	3	2		
71	1	3	3	3	3	4	4	1	
72	1	3	3	5	5	2	2		
73	2	2	1	4	3	2	1		
74	1	3	5	3	1	2	5	1	1
75	1	3	3	3	3	4	3	1	
76	1	3	3	4	5	2	5		
77	1	3	4	4	1	5	4	1	
78	1	2	2	3	3	3	3		
79	1	4	5	4	4	5	4	1	1
80	1	5	5	2	3	4	5	1	
81	1	3	3	3	2	4	5	1	
82	1	3	3	5	1	5	5	1	
83	1	3	3	4	2	5	3	1	
84	1	3	4	4	3	4	3	1	1
85	2	3	3	3	3	4	4	1	1
86	1	4	4	4	2	4	3	1	1
87	1	3	3	3	5	3	4	1	
88	1	3	4	2	3	4	5	1	1
89	1	3	3	2	3	3	3	1	1
90	1	1	1	3	2	3	3	1	
91	1	1	2	1	3	3	3		
92	2	4	1	4	1	1	1		
Question Number:	**Q. 1**	**Q. 2**	**Q. 3**	**Q. 4**	**Q. 5**	**Q. 6**	**Q. 7**	**Q.8A**	**Q. 8B**
Participant									
93	1	2	1	4	2	2	1		1
Sum	**111**	**292.5**	**294**	**302**	**264**	**338**	**338**	**68**	**45**
Average	**1.19**	**3.15**	**3.16**	**3.25**	**2.84**	**3.63**	**3.63**	**0.731**	**0.484**

APPENDIX S

Revised Natural Frequency Tree Method Experiment Complete Statistical Findings and Raw Data

Complete Statistical Findings

Question 1: Do the results from the first test question in the revised natural frequency experiment show significant improvement when compared the first test question results from the original natural frequency tree experiment?

Hypothesis Statements:

Null: The results for test question one from the revised natural frequency tree experiment do not show significant improvement in performance when compared to the results of test question one from the original natural frequency tree experiment.

Alternative: The results for test question one from the revised natural frequency tree experiment do show significant improvement in performance when compared to the results of test question one from the original natural frequency tree experiment.

Assumption Checking:

The assumption of normality is satisfied for the original natural frequency tree experiment as the sample size is considered to be large, i.e. the sample size is at least 30.

The assumption of normality is satisfied for the revised natural frequency tree experiment even though it has small sample size because:

Table A.13: Descriptives Used To Help Determine The Normality Of The Revised Natural Frequency Tree Experiment Sample Population

Descriptives

			Statistic	Std. Error
Question 1	Mean		**.8182**	.12197
	95% Confidence Interval for Mean	Lower Bound	.5464	
		Upper Bound	1.0899	
	5% Trimmed Mean		**.8535**	
	Median		1.0000	
	Variance		.164	
	Std. Deviation		.40452	
	Minimum		.00	
	Maximum		1.00	
	Range		1.00	
	Interquartile Range		.00	
	Skewness		-1.923	.661
	Kurtosis		2.037	1.279

The Mean and 5% Trimmed Mean are approximately equal, which indicates that there are no outliers. This is an important requirement for the normality assumption.

Table A.14: Normality Test Results For The Revised Natural Frequency Tree Method Experiment's Sample Population

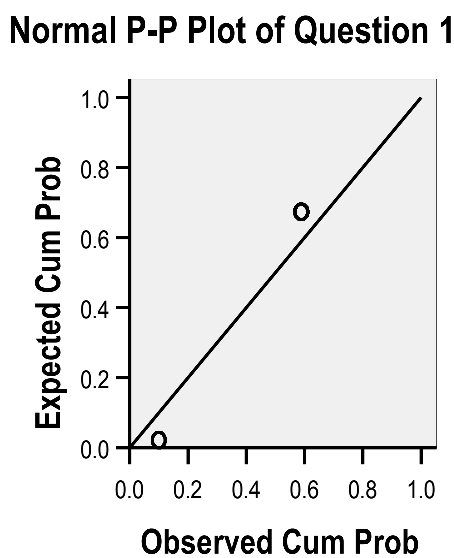
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Question 1	.492	11	**.000**	.486	11	.000

a. Lilliefors Significance Correction

From the above table, the Kolmogorov-Smirnov Normality test shows that (P-value = 0.000) < (α = 0.05), which means that the assumption of normality not satisfied; however, the assumption of normality is robust so it is necessary to also look at the normal probability plot to see if it can be satisfied.

Figure A.3: Normality Plot For The Revised Natural Frequency Tree Method Experiment's Sample Population



Decision is to perform an Independent Two-Samples t-distribution Hypothesis test.

Analysis:

Table A.15: Revised Natural Frequency Tree Experiment And Original Natural Frequency Tree Experiment Statistics

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Question Exp 1	Experiment 1	92	.7609	.42889	.04471
	Experiment 2	11	.8182	.40452	.12197

Table A.16: Independent Samples Test Results For The t-Distribution Independent Two-Hypothesis Test For The Revised Natural Frequency Tree Method Experiment And Original Natural Frequency Tree Method Experiment

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference: Lower	95% Confidence Interval of the Difference: Upper
Question Exp 1	Equal variances assumed	.838	**.362**	-.421	101	.675	-.05731	.13608	-.32725	.21263
	Equal variances not assumed			-.441	12.843	.666	-.05731	.12991	-.33830	.22368

To be able to use the t-distribution independent samples test, one more assumption needs to be checked, which is if the variances are equal. Levene's test does this by using the F-distribution. As seen in the above table the (P-value = 0.362) > (α = 0.05), thus determining that the assumption of equal variances is satisfied. The t-test value and P-value in the "equal variances assumed" row are then used to make the decision about the main question. These values are shown in the table below:

t-test value = -0.421
P-value = 0.675/2 = 0.3375 (The P-value is divided by 2 as a one-tailed test is needed and SPSS only gives a 2-tailed test P-value.)

P-value = 0.3375 and is greater than the level of significance of 0.05, thus we fail to reject the null hypothesis.

<u>Conclusion:</u>

At the 5% level, the results for test question one from the revised natural frequency tree experiment do not show significant improvement in performance when compared to the results of test question one from the original natural frequency tree experiment.

Question 2: Is there a significant difference between the results of the Natural Frequency group (79% baseline) from the natural frequency verification experiment and the results obtained from Test Question 1 in the revised natural frequency tree method experiment?

Hypothesis Statements:

Null: There is no difference between the Natural Frequency group results and Test Question 1 results.

Alternative: There is a significant difference between the Natural Frequency group results and Test Question 1 results.

Assumption Checking:

The assumption of normality is satisfied. See Tables A.13 and A.14 and Figure A.3 above for the normality assumption test results for the revised natural frequency tree experiment's sample population.

Analysis:

Table A.17: Sample Statistics For the First Test Question From The Revised Natural Frequency Tree Method Experiment

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Question 1	11	.8182	.40452	.12197

Table A.18: Complete One-Sample Test Results For The One-Sample t-Distribution Hypothesis Test For The Revised Natural Frequency Tree Method Experiment And Natural Frequency Verification Experiment

One-Sample Test

	Test Value = .79					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Question 1	**.231**	10	**.822**	.02818	-.2436	.2999

t-test value = 0.231
P-value = 0.822
P-value = 0.822 and is greater than the level of significance of 0.05, thus we fail to reject the null hypothesis.

Conclusion:

At the 5% level, the results from Test Question 1 from the revised natural frequency tree experiment are not significantly different than the results obtained from Natural Frequency group in the natural frequency verification experiment.

Raw Data

Table A.19: Correct Number Of Test Question Answers[151] Within 10 Points Of Actual Answer N = 11				
Key: 1 = Answered Question Correctly; Blank = Answered Question Incorrectly				
	Test Question 1	**Test Question 2**	**Test Question 3**	**Test Question 4**
Participant				
1	1	1		
2	1	1	1	
3			1	1
4	1		1	1
5	1	1	1	1
6	1	1	1	1
7	1	1	1	
8			1	1
9	1	1	1	1
10	1	1	1	1
11	1	1	1	1
Total Number Correct:	**9**	**8**	**10**	**8**
Percentage Correct:	**0.818182**	**0.727273**	**0.909091**	**0.727273**

[151] See Appendix I for actual test questions.

APPENDIX T

Trend Line Analysis Complete Statistical Findings

Question: Is there a significant difference between the trend lines created by test questions 1 through 4 from the original natural frequency tree experiment and the revised natural frequency tree experiment?

Hypothesis Statements:

Null: There is not a significant difference between the original natural frequency tree experiment and the revised natural frequency tree experiment trend lines.

Alternative: There is a significant difference between the original natural frequency tree experiment and the revised natural frequency tree experiment trend lines.

Assumption Checking:

The assumption of normality is satisfied for the original natural frequency and revised natural frequency trend lines even though they have small sample sizes (questions 1 through 4) because:

Table A.20: Descriptives Used To Help Determine The Normality Of The Trend Lines Created By The Original And Revised Natural Frequency Tree Method Experiments

Descriptives

	Group			Statistic	Std. Error
Percentage of Correct Answers	Experiment 1	Mean		**.6250**	.05852
		95% Confidence Interval for Mean	Lower Bound	.4388	
			Upper Bound	.8112	
		5% Trimmed Mean		**.6256**	
		Median		.6300	
		Variance		.014	
		Std. Deviation		.11705	
		Minimum		.48	
		Maximum		.76	
		Range		.28	
		Interquartile Range		.23	
		Skewness		-.233	1.014
		Kurtosis		.283	2.619
	Experiment 2	Mean		**.7975**	.04308
		95% Confidence Interval for Mean	Lower Bound	.6604	
			Upper Bound	.9346	
		5% Trimmed Mean		**.7950**	
		Median		.7750	
		Variance		.007	
		Std. Deviation		.08617	
		Minimum		.73	
		Maximum		.91	
		Range		.18	
		Interquartile Range		.16	
		Skewness		.855	1.014
		Kurtosis		-1.289	2.619

The Mean and 5% Trimmed Mean are approximately equal, which indicates that there are no outliers. This is an important requirement for the normality assumption.
The box plot on the next page also supports this.

Figure A.4: Box Plots For The Trend Lines Form The Original And Revised Natural Frequency Tree Method Experiments

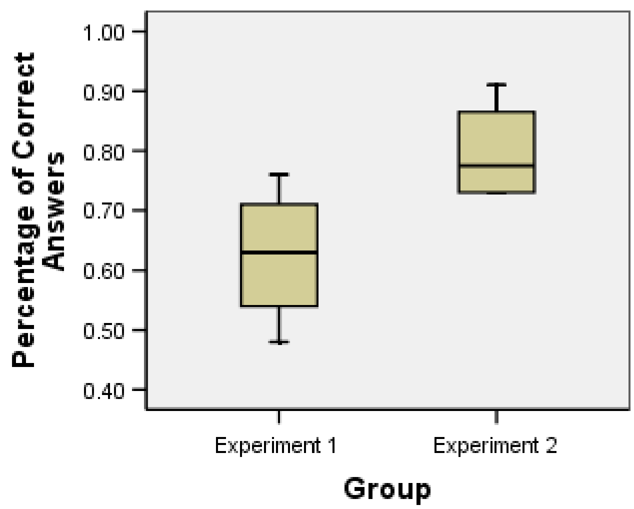


Table A.21: Normality Test Results For The Trend Lines From The Original And Revised Natural Frequency Tree Method Experiments

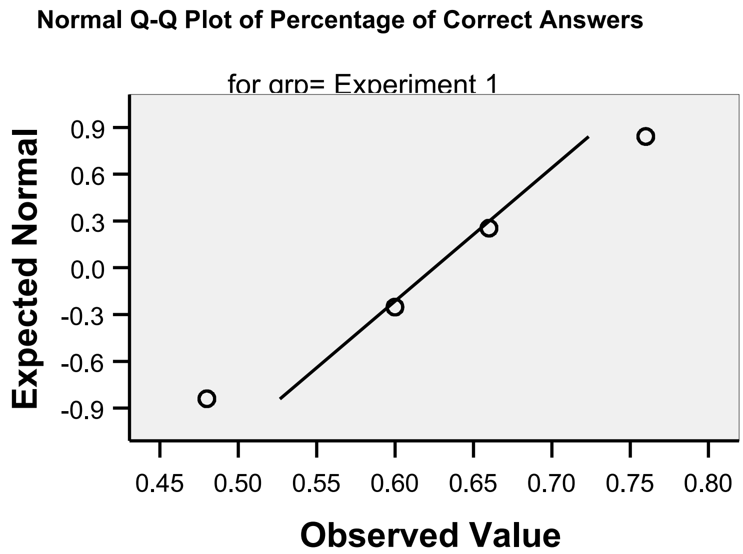
Tests of Normality

	Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Percentage of Correct Answers	Experiment 1	.165	4	.	.997	4	.989
	Experiment 2	.283	4	.	.863	4	.272

a. Lilliefors Significance Correction

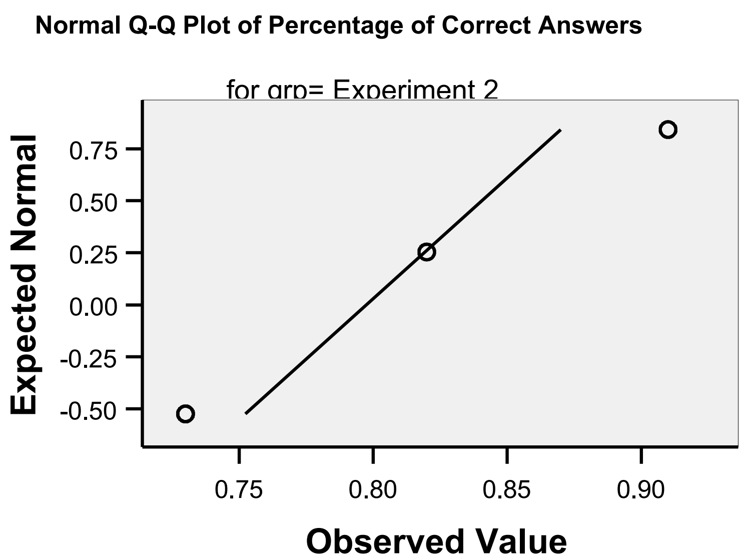
As seen in the above table, the Kolmogorov-Smirnov Normality test does not give any P-values since the sample sizes are very small; however, the assumption of normality is robust, so we need to look at the normal probability plots to see if it can be satisfied.

Figure A.5: Normality Plot For The Original Natural Frequency Tree Method Experiment's Trend Line



Since the points are close to the diagonal line, it is safe to assume that the assumption of normality is satisfied.

Figure A.6: Normality Plot For The Revised Natural Frequency Tree Method Experiment's Trend Line



Since the points are close to the diagonal line, it is safe to assume that the assumption of normality is satisfied.

Decision is to perform an Independent Two-Samples t-distribution Hypothesis test.

Analysis:

Table A.22: Equality Of Variances Assumption For The Original And Revised Natural Frequency Tree Method Experiment Trend Lines

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference: Lower	95% Confidence Interval of the Difference: Upper
Percentage of Correct Answers	Equal variances assumed	.226	**.651**	-2.374	6	.055	-.17250	.07267	-.35032	.00532
	Equal variances not assumed			-2.374	5.514	.059	-.17250	.07267	-.35420	.00920

To be able to use the t-distribution independent samples test, one more assumption needs to be checked, which is if the variances are equal. Levene's test does this by using the F-distribution. As seen in the above table the (P-value = 0.651) > (α = 0.05), thus determining that the assumption of equal variances is satisfied. The t-test value and P-value in the "equal variances assumed" row are then used to make the decision about the main question. These values are shown in the table below:

Table A.23: Complete Trend Lines Independent Samples Test Results For The t-Distribution Independent Two-Hypothesis Test

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference: Lower	95% Confidence Interval of the Difference: Upper
Percentage of Correct Answers	Equal variances assumed	.226	.651	**-2.374**	6	**.055**	-.17250	.07267	-.35032	.00532
	Equal variances not assumed			-2.374	5.514	.059	-.17250	.07267	-.35420	.00920

t-test value = -2.374
P-value = 0.055
P-value = 0.055 and is greater than the level of significance of 0.05, thus we fail to reject null hypothesis.

Conclusion:

At 5% level, there is not a significant difference between the original natural frequency tree experiment and the revised natural frequency tree experiment trend lines.

APPENDIX U

Overall Percentage of Correct Answers from the Original Natural Frequency Tree Experiment and the Revised Natural Frequency Tree Experiment

Table A.24: Percentage Of Correct Answers For The Original Natural Frequency Tree Experiment	
Total Number Questions: 368	Number Of Correct Answers
Question 1	70
Question 2	61
Question 3	55
Question 4	44
Total Correct	230
	230/368
Overall Percentage Correct	**63%**

Table A.25: Percentage Of Correct Answers For The Revised Natural Frequency Tree Experiment	
Total Number Questions: 44	Number Of Correct Answers
Question 1	9
Question 2	8
Question 3	10
Question 4	8
Total Correct	35
	35/44
Overall Percentage Correct	**80%**

APPENDIX V

Revised Natural Frequency Experiment Post-Test Questionnaire Complete Statistical Findings and Raw Data

Complete Statistical Findings

Question: Is there a significant difference between Questionnaire Question 1's average answer and Questionnaire Question 2's average answer?

Hypothesis Statements:

Null: There is no difference between the average answers of Questionnaire Question 1 and Questionnaire Question 2.

Alternative: There is a significant difference between the average answers of Questionnaire Question 1 and Questionnaire Question 2.

Assumption Checking:

The assumption of normality is satisfied for Question 1 and Question 2 as:

Table A.26: Descriptives Used To Help Determine The Normality Of The Sample Populations For The Revised Natural Frequency Tree Method Experiment's Post-Test Questionnaire Questions

Descriptives

			Statistic	Std. Error
Questionnaire Question 1	Mean		**1.2727**	.14084
	95% Confidence Interval for Mean	Lower Bound	.9589	
		Upper Bound	1.5865	
	5% Trimmed Mean		**1.2475**	
	Median		1.0000	
	Variance		.218	
	Std. Deviation		.46710	
	Minimum		1.00	
	Maximum		2.00	
	Range		1.00	
	Interquartile Range		1.00	
	Skewness		1.189	.661
	Kurtosis		-.764	1.279
Questionnaire Question 2	Mean		**3.2727**	.27273
	95% Confidence Interval for Mean	Lower Bound	2.6651	
		Upper Bound	3.8804	
	5% Trimmed Mean		**3.2475**	
	Median		3.0000	
	Variance		.818	
	Std. Deviation		.90453	
	Minimum		2.00	
	Maximum		5.00	
	Range		3.00	
	Interquartile Range		1.00	
	Skewness		.344	.661
	Kurtosis		-.054	1.279

The Mean and 5% Trimmed Mean are approximately equal, which indicates that there are no outliers. This is an important requirement for the normality assumption. The box plots also show that there are no outliers (see next page).

Figure A.7: Box Plot For Revised Natural Frequency Tree Method Post-Test Questionnaire Question 1

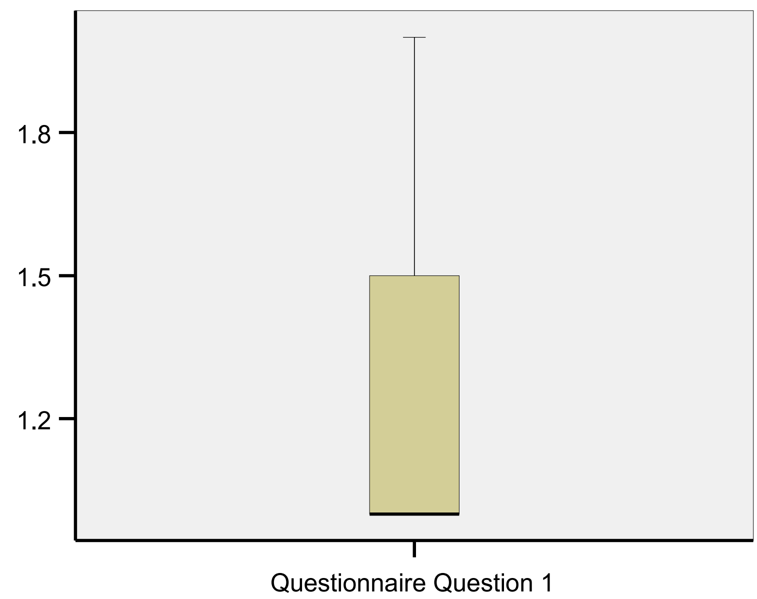


Figure A.8: Box Plot For Revised Natural Frequency Tree Method Post-Test Questionnaire Question 1

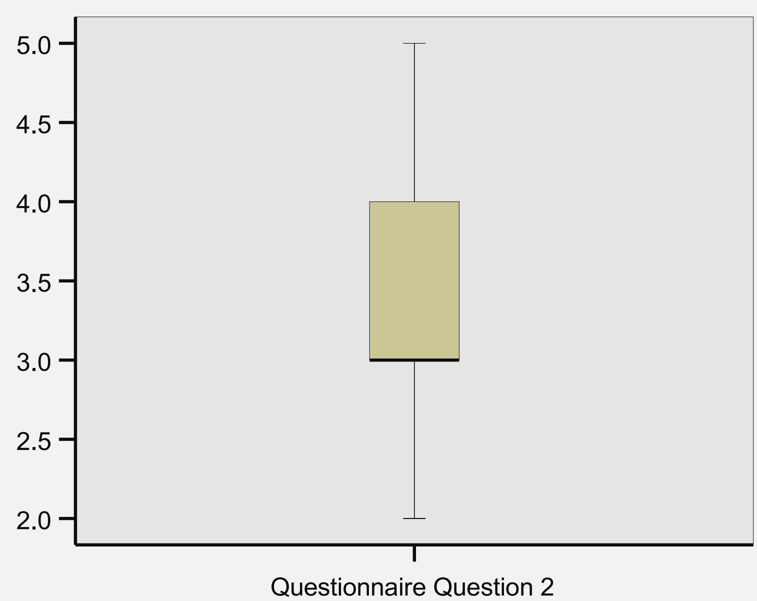


Table A.27: Normality Test Results For The Revised Natural Frequency Tree Experiment Post-Test Questionnaire Questions 1 And 2

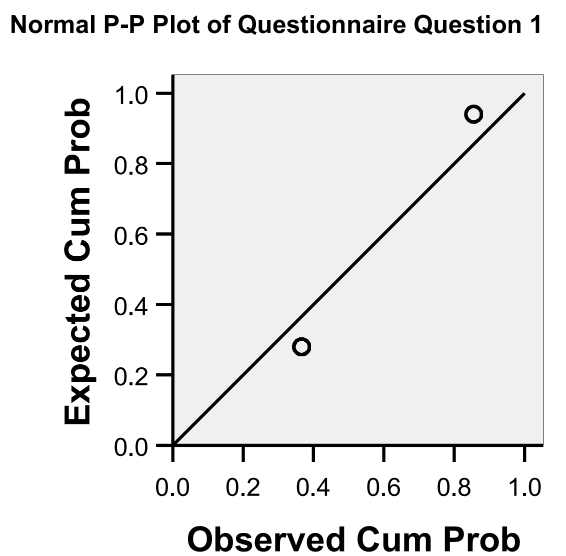
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Questionnaire Question 1	.448	11	**.000**	.572	11	.000
Questionnaire Question 2	.255	11	**.044**	.899	11	.181

a. Lilliefors Significance Correction

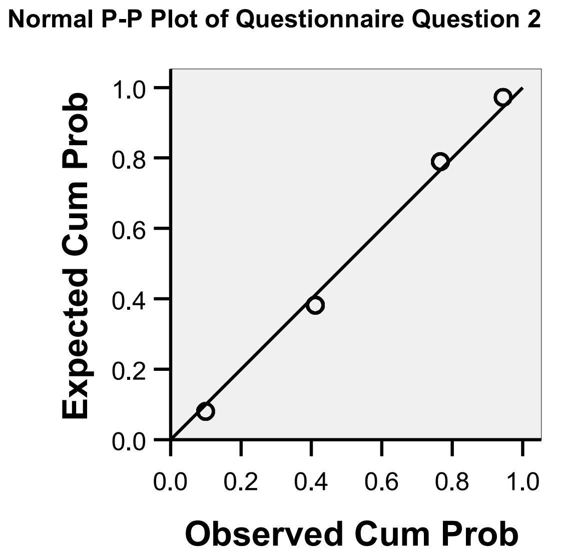
From the above table, the Kolmogorov-Smirnov Normality test shows that (P-value = 0.000) < (α = 0.05), which means that the assumption of normality not satisfied; however, the assumption of normality is robust so it is necessary to also look at the normal probability plots to see if it can be satisfied.

Figure A.9: Normality Plot For Revised Natural Frequency Tree Experiment Post-Test Questionnaire Question 1



Note that the sample only consists of 0 and 1's. Due to this, the plot shows concentrations only at two places. Since they are close to the diagonal line, it is safe to assume that the assumption of normality is satisfied.

Figure A.10: Normality Plot For Revised Natural Frequency Tree Experiment Post-Test Questionnaire Question 1



Since the points are close to the diagonal line, it is safe to assume that the assumption of normality is satisfied.

Decision is to perform a t-test for dependent samples.

Analysis:

Table A.28: Revised Natural Frequency Tree Experiment Post-Test Questionnaire Questions 1 And 2 Sample Statistics

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Questionnaire Question 1	1.2727	11	.46710	.14084
	Questionnaire Question 2	3.2727	11	.90453	.27273

Table A.29: Complete Revised Natural Frequency Tree Method Post-Test Questionnaire Paired Samples Test Results For The t-Test For Paired Samples

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Questionnaire Question 1 - Questionnaire Question 2	-2.00000	1.09545	.33029	-2.73593	-1.26407	**-6.055**	10	**.000**

t-test value = -6.055
P-value = 0.000
P-value = 0.000 and is less than the level of significance of 0.05, thus we reject the null hypothesis

Conclusion:

At the 5% level, the results for Questionnaire Question 1 are significantly different than the results for Questionnaire Question 2.

Raw Data

Table A.30 Answers From Post-Test Questionnaire (Revised Natural Frequency Tree Experiment)									
Key: Questions 1-7: Scale Answers Ranging From 1 - 5 Questions 8A & 8B: 1 = Yes; Blank = No									
Question Number:	**Q. 1**	**Q. 2**	**Q. 3**	**Q. 4**	**Q. 5**	**Q. 6**	**Q. 7**	**Q.8A**	**Q. 8B**
Participant									
1	1	2	3	1	4	3	5		
2	1	3	2	4	2	3	2		
3	1	4	4	3	2	5	5	1	
4	2	4	4	4	3	4	5	1	
5	1	3	4	3	4	3	3	1	1
6	1	3	3	4	2	5	4	1	1
7	2	3	3	2	3	3	3	1	
8	1	4	5	4	2	4	3	1	1
9	1	3	3	3	4	3	5	1	1
10	2	2	3	2	2	3	3	1	
11	1	5	5	4	1	5	4	1	
Sum	**14**	**36**	**39**	**34**	**29**	**41**	**42**	**9**	**4**
Average	**1.27**	**3.27**	**3.55**	**3.09**	**2.64**	**3.73**	**3.82**	**0.82**	**0.36**