SUBJECTIVE PROBABILITIES DERIVED FROM THE PERCEIVED RANDOMNESS OF SEQUENCES OF OUTCOMES

Egan J Chernoff
B.Ed., University of British Columbia, 2001
B.Sc., Thompson Rivers University, 1999

Thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

in the
Faculty of Education

APPROVAL

Name: **Egan J Chernoff**

Degree: **Doctor of Philosophy**

Title of Thesis: **Subjective Probabilities Derived from the Perceived Randomness of Sequences of Outcomes**

Examining Committee:

Chair: Kevin O'Neill, Associate Professor

Rina Zazkis, Professor
Senior Supervisor

Peter Liljedahl, Assistant Professor
Committee Member

Stephen Campbell, Associate Professor
Committee Member

Nathalie Sinclair, Assistant Professor
Internal/External Examiner

Cliff Konold, Associate Professor, University of Massachusetts Amherst
External Examiner

Date Defended/Approved: **June 25, 2009**

ABSTRACT

This research continues the longstanding tradition of taking an interdisciplinary approach to studies in probability education. Respondents are presented with sequences of heads and tails derived from flipping a fair coin five times, and asked to consider their chances of occurrence. A new iteration of the comparative likelihood task, which maintains the ratio of heads to tails in all of the sequences presented, provides unique insight into individuals' perceptions of randomness and associated probabilities. In order to develop the aforementioned insight, this research presents unconventional interpretations of the sample space—organized according to switches, longest run, and switches and longest run, which are all based upon individuals' verbal descriptions of the sample space—to help situate individuals' answers and justifications within conventional probability. In doing so, it will be shown that conventionally incorrect responses to the task are not, necessarily, devoid of correct probabilistic thinking.

The data for this research is based upon two interrelated questionnaires, denoted Study I and Study II. Answers and justifications from the 56 prospective elementary school mathematics teachers in Study I are employed to develop the second iteration of the questionnaire in Study II, which was given to 239 prospective mathematics teachers (comprised of 163 elementary school teachers and 76 secondary school teachers).

To accurately render the data compiled in Study II, an original theoretical framework, entitled the *meta-sample-space*, will be used with a new method, entitled *event-description-alignment*, to demonstrate, for the first time, that individuals' probabilities, derived from the perceived randomness of sequences of outcomes, are in accord with, or model, a *subjective-sample-space* partitioned according to said individuals' interpretation of the sequence of outcomes they are presented.

Keywords:
Abduction; Event-description-alignment; Likelihood; Longest run; Meta-sample-space; Multivalence; Multivalent; Multivalentological; Multivalentology; Outcomes; Partitions; Probability; Randomness; Runs; Sample space; Sequences; Subjective-sample-space; Superposition; Switches.

DEDICATION

To my pack:

To Kristen, your understanding, willingness, and support have made everything that has happened possible. Thank you.

To Scout, for waiting patiently (read: sleeping) at my feet when I needed to work, and for needing a walk, rain or snow, when I needed a break.

ACKNOWLEDGEMENTS

Before I get to positive acknowledgements, I want to document the incessant frustration derived from Microsoft Office 2008 for Mac. More specifically, Office 2008's incompatibility with the Spaces feature in Mac OS X Leopard has, without a doubt, been the bane of my disquisition.

With that out of the way, I want to first acknowledge the unconditional support from my family: Jack, Nina, and Kevin. The importance of education has always been an implicit assumption in our family, and for that I am grateful.

Second, I want to acknowledge three friends Dave Cacchioni, Doug Pope, and Aaron Mueller for—and surpassed only by Kristen in—their support, encouragement, frequent liquid therapy sessions, and (especially) comic relief during my graduate studies. Dave, Tuesday afternoons/evenings have not been, and never will be, the same. Don't forget to keep up the in depth studies on your favorite type of ellipse! By the way, did you ever finish that poster board? Aaron, thank you for always being interested in my research and asking questions, as only a true *cybrarian* would. Doug, thanks for constantly, and I mean constantly, referring to your published articles—especially the one in the *Malaysian Journal of Chemistry*—as a reminder about the importance of publish or perish.

Third, I would like to acknowledge all of my classmates in the Secondary Mathematics Education Master's Program and the Mathematics Education Doctoral Program at Simon Fraser University. However, special thanks are due to Craig Newell. Sure, Craig was one of our classmates, but, to be honest, he was not your conventional classmate. Craig was one of those unique individuals who remembered everything he had ever read, and had read nearly everything. Craig's experience, knowledge, and, most importantly, willingness to listen provided his "classmates" with the rare opportunity of taking doctoral courses with two mathematics education academicians in the room. I was one of those people fortunate enough to have experienced Craig's wisdom while taking my doctoral classes, and to have been able to thank him personally for the dual role he played.

Fourth, I must recognize the academic triad (read: Triple Deity) of Rina, Sen, and Peter for their instruction both in and out of class, and introducing me to the North American and International mathematics education communities, undeservedly, as a colleague. Further, thanks to Rina (and Peter) for making it feasible to not only be a graduate student, but a graduate student who travelled the world (Hola!). Speaking of finances, thank you to the Rotary Club of Burnaby, the Lis Welch Endowment Fund, and the Social Sciences and Humanities Research Council for helping fund my research.

Finally, but most importantly, I would like recognize my senior supervisor Professor Rina Zazkis. While there are innumerable reasons to detail at this point, Professor Zazkis would not allow it. As such: Thank you Rina. We have come a long way from our first verbal exchange in Math 603. Good luck with the band!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

At the very core of my research is subjective probability. In fact, I see my research as the culmination of an intense interest with subjective probability. While certain mathematics education researchers claim, at present, that cognitive research on the subjective approach to probability does not exist, I disagree. In fact, I would argue that mathematics education literature and psychology literature are saturated with cognitive research on the subjective approach to probability. For example, I would argue that the normative, heuristic, and informal approaches to probability all fall under the category of cognitive research on subjective probability. In other words, because research has not been conducted specifically on the degree to which an individual believes in a particular proposition, does not mean there is an absence of cognitive research on subjective probability. Influencing the difference between my opinion and the opinion of others is how one defines subjective probability. Complicating matters, subjective probability means different things to different individuals. Alternatively stated, subjective probability is subjective. As such, and for contextual purposes, I present certain research that has influenced my interpretation of the present state of subjective probability in the field of mathematics education.

I contend the lack of a unified definition for subjective probability best captures and influences the current state of subjective probability in mathematics education. Further, the definitional issues that exist in mathematics education also exist in probability theory. In fact, the issues seen in mathematics education are, for the most part, derived from issues in probability theory. Gillies (2000) states:

> The difficulty with this terminology is that the 'subjective' interpretations of probability include both the subjective theory of probability, which identifies probability with degree of belief, and the logical theory, which identifies probability with degree of rational belief. Thus, subjective is used both as a general classifier and for a specific theory. This is surely unsatisfactory (p. 19).

To address the difficulty presented, probabilistic philosophers have made further distinctions within subjective probability. For example, Hacking (2001) notes a distinction between "personal" probabilities and "interpersonal" (p. 32) probabilities. Unfortunately, the distinction between personal and interpersonal within subjective probability does not have a counterpart in mathematics education, yet. As such, and at present, subjective probability aligns with both the personal and interpersonal theories, and the lack of counterparts is representative of the current state of subjective probability measurement in mathematics education.

The state of subjective probability in mathematics education is also influenced by philosophical underpinnings. "The research of psychologists Daniel Kahneman and Amos Tversky…has provided mathematics educators with a theoretical framework for researching learning in probability" (Shaughnessy, 1992, p. 470). Kahneman and Tversky's use of subjective probability is as follows:

> We use the term "subjective probability" to denote any estimate of the probability of an event, which is given by a subject, or inferred from his behavior. These estimates are not assumed to satisfy any axioms of consistency requirements. We use the term "objective probability" to denote values calculated, on the basis of stated assumptions, according to the laws of the probability calculus. It should be evident that this terminology is noncommittal with respect to any philosophical view of probability (Tversky & Kahneman, 1982, p. 32).

Despite the noncommittal stance from (arguably and with all due respect to Piaget and Inhelder) the fathers of probability education and the polysemic nature of the term subjective probability, the current state of subjective probability in mathematics education is, I contend, aligned with the personal interpretation of subjective probability. In an attempt to demonstrate my assertion, definitions for subjective probability in mathematics education from significant pieces of literature are presented:

- "*subjective probability*, describes probability *as* a degree of belief, based upon personal judgment and information about an outcome" (Jones, Langrall, & Mooney, 2007, p. 913);
- "In this subjective view, what is random for one person might be nonrandom for another…a degree of uncertainty specific for each person" (Batanero, Henry, & Parzysz, 2005, p. 24);
- "The basic assumption here is that subjects have their own probabilities which are derived from an implicit preference pattern between decisions" (Borovcnik, Bentz, & Kapadia, 1991, p. 41);
- "To a greater or lesser extent the probability is an expression of personal belief or perception" (Hawkins, & Kapadia, 1984, p. 349).

The examination of certain subjective probability definitions evidences an implicit(ly circular) definition of the term subjective. Subjective probability, when used in the field of mathematics education, is implicitly aligned with (1) the subjective theory of probability and not as a general classifier, and (2) the personal theory of subjective probability, not the interpersonal theory.

Given the implicit state of subjective probability in mathematics education described above, I contend there does exist a variety of cognitive research on the subjective approach to probability. In fact, I would argue that in my impending review of the literature, all research presented pertains to cognitive research on subjective probability. Further, I would also argue that the research I am about to present contributes to the existing cognitive research on the subjective approach to probability. Again, and as mentioned, other individuals may disagree with the statements I make here, but, and as I will soon show, who is considered ‘correct’ is a matter of interpretation.

Egan J Chernoff
January 26, 2009

INTRODUCTION

"The shooting of the hunters was dreadful" (Paulos, 1980, p.65). A number of interpretations coexist for the aforementioned statement. For example, if the hunters do not work on their aim they will never be able to hit their intended target, i.e., the shooting of the hunters was dreadful; in this particular interpretation of the statement the hunter's shooting ability is called into question. For another example, what did that cute, defenseless baby animal ever do to deserve being shot between the eyes, i.e., the shooting of the hunters was dreadful; in this second interpretation the shooting of an animal by the hunters is considered dreadful and the hunter's shooting ability is not under consideration. For yet another example, it should be understood that when people are walking around the woods carrying loaded weapons accidents are bound to happen, i.e., the shooting of the hunters was dreadful; in this third interpretation it is the hunters who are shot, not the animals, and (arguably) the hunters' shooting ability is not taken into consideration. There (co)exist at least three possible interpretations of the statement "the shooting of the hunters was dreadful;" at least three because the "shooting" discussed previously was conducted with some type of firearm and not a camera. As demonstrated, the statement is *multivalent* (i.e., has many distinct interpretations).

There is concurrency associated with the multiple interpretations of a multivalent statement. One method to describe the coexistence of interpretations is to declare the statement exists in a state of *superposition*, i.e., the statement represents all possible interpretations whether enumerable or not: poor aim, dead animals, dead hunters, and photographic shooting. Moreover, when the statement is in a state of superposition there is no one particular interpretation for the statement, because all interpretations exist at once. Limiting the statement to a single possibility, or to *collapse* the state of superposition, requires a particular interpretation to take place. Interpreting the statement “the shooting of the hunters was dreadful” collapses the state of superposition, i.e., coexistence of all possible interpretations, and limits the statement to one particular interpretation. For example, one individual’s interpretation may result in a hunting accident interpretation, whereas another individual’s interpretation may result in a marksman interpretation.

To determine which particular collapse of the superposition of interpretations has taken place, inferences can be made through the examination of comments made by individuals who have read the statement. Consider, for example, an individual who after reading the statement comments, “I guess even if you wear a gaudy orange vest that does not mean you are immune from accidents.” One may reasonably infer that it is more likely that the individual has interpreted the statement in terms of a hunting accident interpretation, as opposed to any of the other available interpretations. Further, and as another example, the reading of an individual’s comments such as, “they need to spend more time practicing before they go out into the woods” causes one to reasonably infer that it is more likely the individual interpreted the statement as a hunting accuracy interpretation, rather than any of the other available interpretations.

Examination of comments not only provides insight into the collapse of the superposition of interpretations for a particular individual, but also provides the opportunity for determination of whether an individual’s interpretation matches the intended interpretation of an author. If a third party is the creator of the statement, or knows with certainty the intended interpretation of the author, they are then able to, by the reading of comments made by an individual, determine if the individual's interpretation aligns with the author's interpretation. Consider, for example, if it is known through some manner that the author who wrote the statement was in fact discussing hunting accidents. The individual who commented, "they need to spend more time practicing before they go out into the woods" has an interpretation of the statement that is less likely to align with the intended interpretation of the author. The individual who commented, "I guess even if you wear a gaudy orange vest that does not mean you are immune from accidents" is more likely to be aligned with the intended interpretation of the author.

The research method presented—consisting of: statement, multivalence, superposition, interpretation, collapse, comment, inferences, and intended interpretation—does not change throughout the remainder of the research; however, *what* is analysed via the research method does change. In general, the research method, as just detailed, can be described as a quantum mechanical exploration, examination, critique, creation, and testing of hypotheses generated via inference when individuals engage with multivalence. The most efficient way to describe the research method is to build on the already-defined terms of multivalent and multivalence to define: *multivalentology* as (a) the study of multivalence, or (b) the quantum mechanical exploration, examination, critique, creation, and testing of (i.e., the study of) hypotheses generated via inference when individuals engage with multivalence; *multivalentological* as pertaining to multivalentology; and *multivalentologist* as a person who studies multivalence. Multivalentology is not restricted to statements. For example, the impending research is described by the author—a self-proclaimed multivalentologist—as a multivalentological disquisition on subjective probabilities involving the perceived randomness of sequences of outcomes.

An oft-used task found in psychology and mathematics education (an example is shown in Figure 1) will act as the medium for investigation. Despite a general structure (e.g., binomial experiment, probability of success equaling probability of failure, two or more sequences presented) associated with different variations of the task, there does not exist a common name for the task (e.g., HT-sequence problem, sequence task). As such, the author wishes to denote the task presented below (akin to different variations found in the literature) as the Comparative Likelihood Task, hereafter referred to as the CLT.

Which of the following sequences is the least likely to occur from flipping a fair coin five times. Justify your response:

a) TTTHT
b) THHTH
c) HTHTH
d) HHHTT
e) all four sequences are equally likely to occur

Figure 1. The comparative likelihood task: An example

The first chapter is dedicated to a review of the literature on the CLT. Recognizing the influential role psychologists have played in research on the CLT and probability education in general, Chapter One begins with a review of the seminal work on heuristics and biases by Amos Tversky and Daniel Kahneman. Next, literature in mathematics education is reviewed according to chronological periods of probability research in mathematics education.

In Chapter Two the author's novel framework based on the multivalence (i.e., characteristic of having many distinct interpretations) of CLT responses is presented. With the new *CLT Response Interpretation Framework (RIF)* in mind and embracing the notion that a literature review is, in essence, the explication of one particular or multiple interpretations, the Chapter Two concludes with an interpretation of the research literature;

achieved through the exploration, examination, and critique of hypotheses generated from research on the CLT found in Chapter One.

The last half of Chapter Three is comprised of the analysis of results from Study I, given to 56 prospective elementary school mathematics teachers. While results are first analyzed via the multivalence of the CLT Response Interpretation Framework (RIF), results are further analyzed through the author's second framework, the *CLT Task Interpretation Framework (TIF)*, based on the multivalence of the comparative likelihood task, which is developed at the beginning of Chapter Three.

With the interpretive nature of the research in mind, Chapter Four presents the raw data from Study II, where 239 prospective mathematics teachers—comprised of 163 elementary school teachers and 76 secondary school teachers—are presented with an evolved version of the CLT implemented in Study I. However, it is not until Chapter Six—after the development of (1) the notion of a *subjective-sample-space*, (2) a framework known as the *meta-sample-space*, and (3) a description of the method, entitled *event-description-alignment*, in Chapter Five—that the results shown in Chapter Four are analysed.

While conclusions are also presented, Chapter Seven, the final chapter, is dedicated to a discussion on past, present, and future studies that investigate probabilities associated with the perceived randomness of sequences of outcomes. Finally, and as is customary, a research agenda for the CLT and the explicit statement of contributions to research in mathematics education are presented.

In general, the research described is derived from two main goals of the author. First, demonstrate the multivalence of elements of the CLT. Second, demonstrate that certain individual's answers to the comparative likelihood task accord to a subjective organization of the sample space, which is based on their interpretations of sequences of outcomes. Throughout the research described above, it is shown that unconventional views of the sample space—organized according to constructs referred to in what follows as switches, longest run, and switches and longest run, which are all based upon individuals' verbal descriptions of the sample space—can help situate individuals' answers and justifications within conventional probability. In doing so, it is shown that normatively incorrect responses to the task are not, necessarily, devoid of correct probabilistic ways of thinking and, in fact, model particular partitions of the sample space. To aid in explanation the author proposes an original theoretical framework, entitled the meta-sample-space, which will be used with a new method, entitled event-description-alignment, to demonstrate, for the first time, that individuals' subjective probabilities involving perceptions of randomness for sequences of outcomes are in accord with, or model, a subjective-sample-space.

CHAPTER ONE

A Review of the Literature

This research continues the longstanding tradition of taking an interdisciplinary approach to studies in probability education. Analyses and interpretations of CLT responses, which—as will be shown—originated in psychology and have subsequently been used in mathematics education, will draw upon knowledge derived from research in both fields.

Psychology Literature

According to Shaughnessy (1992), probability research—past and present—in mathematics education has been influenced by psychologists Amos Tversky and Daniel Kahneman. As such, the review of the literature begins by focusing on the research of "a key psychological contribution" (Jones & Thornton, 2005, p. 73) to probability education.

Judgment under uncertainty: Heuristics and biases

Central to Tversky and Kahneman's initial research was the notion of *judgment under uncertainty*. Tversky and Kahneman (1974) raised the question: "How do people assess the probability of an uncertain event or the value of an uncertain quality" (p. 1124)? Whether referred to as judgment under uncertainty or intuitive judgments of probability (Tversky & Kahneman, 1974), subjective probabilities (Kahneman & Tversky, 1972), or probability estimates (Jones & Thornton, 2005; Kahneman & Tversky, 1972), "perhaps the most general conclusion, obtained from numerous investigations, is that people do not follow the principles of probability theory in judging the likelihood of uncertain events" (Kahneman & Tversky, 1972, p. 430). More specifically, Tversky and Kahneman (1974) found that "people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operation" (p. 1124).

While the "*heuristics and biases approach* began with a survey of 84 participants at the 1969 meetings of the Mathematical Psychology Society and the American Psychological Association" (Kahneman & Frederick, 2002, p.49), the initial body of research pertaining to the heuristics and biases program quickly grew (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1971, 1973a, 1973b, 1974, 1980, 1982a, 1982b, 1982c, 1983), and now appears to know no bounds (e.g., Kahneman, Slovic, & Tversky, 1982, Gilovich, Griffin, & Kahneman, 2002).

Tversky and Kahneman (1974) described three heuristics (i.e., simpler judgmental operations)—representativeness, availability, and adjustment and anchoring—that are used in the subjective assessment of probabilities (i.e., probability estimates). Furthermore, the authors described particular biases associated with employment of each of the aforementioned heuristics. While a number of diverse tasks were used in the research, it became apparent that particular tasks were associated with particular heuristics. For example, "One oft-used task that evokes representativeness asks subjects to compare likelihoods of sequences of outcomes that have been generated by a binomial process" (Shaughnessy, 2003, p. 219). In recognition of this point, and in recognition of the current scope of the review of the literature (i.e., a focus on representativeness), research on the representativeness heuristic and its related biases is now examined in further detail.

Representativeness heuristic

In examining how "people replace the laws of chance by heuristics" (Kahneman & Tversky, 1972, p. 430), Kahneman and Tversky produced an initial investigation into the representativeness heuristic. According to their findings, an individual who follows the representativeness heuristic "evaluates the probability of an uncertain event, or a sample, by the degree to which it is: (*i*) similar in essential properties to its parent population; and (*ii*) reflects the salient features of the process by which it is generated" (p. 431). Alternatively stated, and more specifically related to a task such as the CLT, the determinants of representativeness were broken down into two particular features: similarities between the sample and its parent population and apparent or perceived randomness. Kahneman and Tversky theorized that events are considered more probable when appearing more representative, and, similarly, events are considered less probable when appearing less representative. In order to test their theory, the authors focused on some now well-known tasks.

Kahneman and Tversky (1972) presented individuals with birth sequences that were considered equally likely, but were hypothesized by the authors to not be "equally representative" (p. 432). Of the three sequences presented—GBGBBG, BGBBBB and BBBGGG—the sequence BGBBBB was considered less likely than GBGBBG because BGBBBB does not reflect the ratio of boys to girls found in the parent population. Further, BBBGGG was deemed less likely than GBGBBG because BBBGGG did not reflect the random nature associated with the birthing of boys and girls in a family. While the findings supported the authors' initial hypotheses, i.e., more representative sequences would be judged more likely, they also declared that similarity of a sample to a parent population is a necessary but not sufficient determinant of representativeness. As such, they further investigated the reflection of randomness determinant.

Local representativeness. In an investigation into the appearance of randomness, Kahneman and Tversky (1972) declared that "two general properties, irregularity and local representativeness, seem to capture the intuitive notion of randomness" (p. 433). To explicate their point, the authors showed how alternating sequences, e.g., a perfect alteration of heads and tails, was too regular and thus would not correspond to the result of a random process (i.e., a sequence such as HTHTH would be too regular to be considered or perceived as random). Local representativeness, on the other hand, is the belief that "the essential characteristics of the parent population are represented not only globally in the entire sample, but also locally in each of its parts" (p. 434). For example, individuals when examining a short sequence of coin tosses would expect (1) the ratio of heads to tails to be close to one, and (2) short runs, which would correspond to frequent alterations, because, as the authors indicated, "People view chance as unpredictable but essentially fair" (p. 434). Of note, the features presented above are associated with the perception of randomness, not randomness.

Further investigation into local representativeness on the part of Tversky and Kahneman (1971) allowed them to "characterize the expectancy of local representativeness" (p. 434) as the "law of small numbers, which asserts that the law of large numbers applies to small numbers as well" (p. 107). While the law of large numbers does assert that a large enough sample taken from a population would be representative of the population, the same cannot be said for small samples. Thus, applying the law of large numbers to a small sample is mistaken, and doing so results in many ill-conceived notions regarding randomness. In a later article, Tversky and Kahneman (1974) began to categorize the biases associated with the representativeness heuristic and entitled them the *misconceptions of chance* (p. 1128) biases, which were associated with (1) the perception of random processes, and (2) the ratio of heads to tails of short sequences.

Representativeness determinants

As witnessed, there are subtleties associated with employing the representativeness heuristic as the "likelihood of an event is evaluated by the degree to which it is representative of the major characteristics of the process or population from which it originated" (Kahneman & Tversky, 1972, p. 452). In order to help arrange these subtleties a determinant distinction diagram, created by the author, is found in Figure 2.

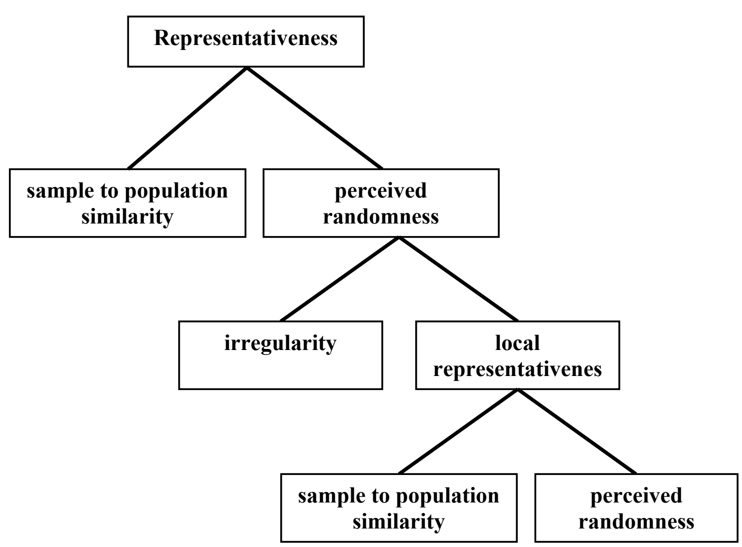

Figure 2. Organization of representativeness determinants

As shown in Figure 2, the major characteristics of the *population* from which a sample originated are represented by *sample to population similarity*, while major characteristics of the *process* from which the sample originated is represented by *perceived randomness*—the first distinction. However, *perceived randomness* itself is further categorized into two determinants as well: *irregularity* and *local representativeness*. Analogous to *sample to population similarity*, *irregularity* has no further distinctions. Analogous to *representativeness* and *perceived randomness*, *local representativeness* also possesses a further determinant distinction: *sample to population similarity* and *perceived randomness*. However, for *local representativeness* the major characteristics of *process* and *population* stem from individuals erroneous intuitions that chance is *unpredictable* and *fair*, respectively. While the authors hypothesized that representativeness plays a role in calculating uncertainty in a variety of situations, the examples discussed and presented in the research, such as the CLT, were of the type that were able to be solved using the mathematics of probability. Of importance to tasks such as the CLT are (1) irregularity, and (2) local representativeness which is broken into (2a) sample to population similarity and (2b) perceived randomness.

Mathematics education research

There is no denying "the research of psychologists Daniel Kahneman and Amos Tversky, and many of their colleagues, has provided mathematics educators with a theoretical framework for researching learning in probability" (Shaughnessy, 1992, p. 470). Further, while "there is little doubt of the importance of their perspective for diagnosing the psychological bases of subjects' misconceptions of probability and statistics" (p. 470), researchers in mathematics education have subsequently built upon the theoretical frameworks from psychology to reflect their interpretations of responses to the CLT. In an attempt to provide order to how similar CLT responses are interpreted differently by different researchers, an observation of the literature found in mathematics education is now conducted.

In their historical synopsis of research on the teaching and learning of probability, Jones and Thornton (2005) classified research into three chronological periods “designated as follows: Phase 1, the Piagetian Period; Phase 2, the Post-Piagetian Period; and Phase 3, the Contemporary Period” (p. 66). In an analogous fashion, the review of mathematics education research literature (pertaining to investigating and interpreting responses to the CLT) will also be reviewed in a chronological manner. First, adopting Jones and Thornton periods, and then adopting Artefactual Period (Chernoff, 2008) to denote the period following the Contemporary Period. However, the current review does not necessarily begin in the Piagetian Period, because research on the CLT made its foray into the field of probability education during the Post-Piagetian period of the 1970s and 1980s.

The Post-Piagetian Period

Jones and Thornton (2005) characterize the Post-Piagetian Period as one of “continuing research of psychologists and the burgeoning growth of studies by mathematics educators” (p. 70). More specific to the CLT, the Post-Piagetian Period in mathematics education saw the continuation of Tversky and Kahneman’s work on the CLT conducted by researchers (Green, 1983; Falk, 1981; Shaughnessy, 1977; 1981) outside the field of psychology, and thus from a different perspective. Shaughnessy (1983, 1992) characterized differences between researchers from psychology and mathematics education: Researchers from psychology were considered "observers and describers" (1992, p. 469), while researchers from mathematics education were considered natural "interveners" (1992, p. 469). The notion of intervener was initially in reference to how mathematics educators felt compelled to help improve students' knowledge, but intervener, as will be seen, as a descriptor for researchers in mathematics education is appropriate on more than one level.

Shaughnessy's (1977, 1981) work with the CLT bridged, in a metaphorical sense, the gap between the fields of mathematics education and psychology. His work went beyond reporting psychological research findings to the mathematics education community. Shaughnessy's research on the CLT addressed some of the inherent inferential issues of Tversky and Kahneman's work and, as such, furthered research involving the CLT as well as the task itself. As Shaughnessy (1992) reflected: "There was no attempt made [by Tversky and Kahneman] to probe the thinking of any of these subjects" (p. 473). To contextualize the aforementioned quote, Shaughnessy's (1977) research, introduced two new, important elements to the task. First, in comparing the chances of occurrence of different sequences, Shaughnessy's task iteration gave students the option of choosing "about the same chance" (p. 309) as one of the forced response items. Second, Shaughnessy asked participants to supply a reason or justify their responses. "In this way it was possible to gain [deeper] insight into the thinking process of the subjects as they answered the questions" (p. 308).

With the "equally likely" and "supply a reason" components added to the task, Shaughnessy provided the opportunity for researchers in mathematics education to begin to depart from the psychological interpretations of CLT responses. Shaughnessy's new components of the task not only allowed for new areas of investigation (i.e., normatively correct answers based on incorrect reasoning and normatively incorrect answers based on correct reasoning), but the elements of the task also allowed for reinforcing research results that were once inferred from forced response items.

Shaughnessy (1977, 1981) determined that, in some instances, subjects chose that all sequences were equally likely to occur, but were equally likely to occur for the individual because each outcome had the same number of boys and girls (i.e., normatively correct answer based on incorrect reasoning). More specifically, the sequence of births BGGBGB and BBBGGG were considered to equally likely because they had the same number (three) of boys and girls. Despite any differences, the main results of Shaughnessy's research echoed the research results of Tversky and Kahneman. More specifically, Shaughnessy's (1981) research with eighty college undergraduate students concluded that an "overwhelming number of students chose the sequence BGGBGB to be more likely to occur" (p. 91) than BBBBGB, and BGGBGB more likely than BBBGGG. With the new "supply a reason" component of the task, Shaughnessy was able to determine that subjects did in fact find the sequence BBBGGG not representative of randomness, and the sequence BBBBGB was not a representative ratio of boys to girls. With students able to supply a reason for their choice, Shaughnessy's work concurred with what Tversky and Kahneman had inferred from their analysis of forced response items.

Shaughnessy's (1977, 1981) research also triggered a focal shift in the interpretation of responses for subsequent research on the CLT in mathematics education during the Post-Piagetian Period. Further numerical evidence that subjects would choose one sequence to be more or less likely than another was no longer the main interest because attention from certain researchers (Falk, 1981; Green, 1983) turned to the interpretations of subjects responses of the CLT. Falk (1981), who recognized the multivalence inherent to CLT responses, noted “the claim that subjects’ deviations from their distorted image of randomness may be true; however, there is an inferential leap in drawing that conclusion. Failure to perceive randomness is but one possible explanation” (p. 223). Other research began to investigate the connection between answer and justification. Rubel (2007), “Green’s (1983) report provided more detail about students’ justifications of responses...and also pointed out how some of these strategies generate correct answers under certain circumstances and incorrect answers under others” (p. 534). Of note, while the justification of the answer to the task (not the answer alone) was primarily used to interpret CLT responses, inferences were still being made, as they were in the psychological research. The difference: inferences were being made on responses that correlate to subjects’ justification for their answer and not on the subjects’ answer to the question.

Transition: Post-Piagetian to Contemporary

Because of the new components added to the CLT, research in the Post-Piagetian Period was characterized by a transition. Instead of inferences being made on answers, inferences were being made on response justifications. Similarly, research in the transition period between Post-Piagetian and Contemporary was also characterized by a transition. Instead of inferences being made on responses and justifications, inferences were made on how respondents interpreted the task. However, the transition period was also characterized by several investigations into perceived randomness.

Perceived randomness. The notion of randomness, more specifically perceived randomness, received much attention from researchers in both mathematics education and psychology during the transition from the Post-Piagetian Period to the Contemporary Period. Interestingly, a number of researchers (Batanero & Serrano, 1999; Batanero, Green & Serrano, 1998; Bennett, 1998; Falk, 1981; Green, 1983; Lecoutre, 1992; Lecoutre, Rovira, Lecoutre, & Poitenineau, 2006) suggest that humans are poor judges of randomness. Reasons for this inability to perceive randomness can *now* be expected as "there are still controversies over the interpretation of basic concepts, such as the meaning of probability and independence" (Batanero, Green & Serrano, 1998, p. 113). Further, and according to LeCoutre et al. (2006), "individuals hold a wide range of meanings for the concept of randomness" (p. 30). Claims concerning randomness and perceptions of randomness should be made with great trepidation. Furthermore, claims made from perceiving perceptions of randomness—as is done with research involving the CLT—should be made with the greatest of trepidation. In recognition of this point, some researchers (Batanero, Green & Serrano, 1998, Batanero & Serrano, 1999; Falk, 1981, Falk & Konold, 1997; Green, 1983, 1988; Lecoutre, 1992; Schilling, 1990; Toohey, 1995) who recognize that the evaluation of randomness is a vital component in research on the sequence task have focused their research on perceived randomness strictly to sequences of outcomes.

Examining interpretations of the task's sequences of outcomes via the appearance of randomness, Falk's (1981) research raised a subtle, yet important, distinction regarding the subjects' ability to judge randomness: not only did the subjects perceive randomness when randomness was not present, subjects also perceived a lack of randomness when randomness was present. This distinction resulted from research where Falk presented students with two sequences and asked which of the sequences appeared more random. Results showed that perception of randomness for the task presented focused upon the notions of switches and runs. More specifically, the results showed that sequences in which more switches occurred appeared more random and sequences in which a long run occurred appeared to be less random. As Batanero and Serrano (1999) found, students' ability to determine what was considered random for results from flipping a fair coin was derived from "the lengths of the runs and, consequently, the proportion of alterations" (p. 560). In other words, randomness was perceived through frequent switches and thus short runs; the caveat is that a sequence with longer runs is more likely to be random (Shaughnessy, 1992, 2003).

Interpreting task interpretations. Research on the CLT during the transition from the Post-Piagetian period to the Contemporary period (Borovcnik & Bentz, 1991; Garfield & Ahlgren, 1988) is also characterized by researchers' interpretations of participants' interpretations of the task, that is, researchers began to question the question. For example, Garfield & Ahlgren (1988) found, "an important factor in misjudgment is a misperception of the question being asked" (p. 55).

Borovcnik and Bentz (1991) embraced and explored the multivalence of the CLT when they questioned the overarching assumption that "items represent a situation which allows for one intuition and a *unique* solution so that certain answers and the reasoning illustrate the understanding of concepts" (p. 76). In order to explicate their point and thus demonstrate their belief that the one-intuition/unique-solution approach is flawed, Borovcnik and Bentz explored a number of probability questions (e.g., tossing a counter, hat lottery, dependent urns) by (1) examining results, (2) interpreting results, and (3) "discuss[ing] various potential strategies to reveal the complexity of the item and of interpreting the answers" (p. 78). In "The six children problem" (p. 84)—a variation of the CLT using B for boys and G girls—Borovcnik and Bentz extended the results of Shaughnessy's (1981) research on the task. Borovcnik and Bentz's analysis of why BGGBGB was more representative of the relative frequency of the birthing process was followed by four potential strategies, which would show how a subject could "end up with a different but appropriate solution" (p. 76). For example, "The answer (C) 'same chance for each' may have been confused with a probability ½ for each sequence" (p. 85). They also demonstrated that "another source of confusion lies in the wording" (p. 85). From this confusion Borovcnik and Bentz detailed how the two sequences presented (BBBBGB and BBBGGG) could be interpreted as one sequence with three boys and three girls and one sequence with five boys and one girl. Moreover, "if subjects base their answers on classes of sequences, i.e., the number of boys, and not on single sequences, they are right" (p. 85). More specifically, the subjects are incorrect in the sense that each of the sequences is equally likely to occur, but if the question is being interpreted as which number of boys in a family of six is more likely, the subjects are correct to say that having three boys (20/64) is more likely than five boys (6/64).

While psychologists initially focused their research on interpreting answers to the CLT and mathematics educators in the Post-Piagetian Period shifted their focus to interpreting responses provided by subjects, the transition period from the Post-Piagetian Period to the Contemporary Period is characterized by a shift in focus towards interpreting subjects' interpretations of the task. While all the focuses can be seen as different, there is one constant to all investigations: the use of inferences in the researchers' interpretations of the CLT responses in an attempt to account for subjects' seemingly inconsistent responses to the task (i.e., mathematics educators as interveners). Hypothesizing about these inconsistent responses defines the Contemporary Period of research in probability education, yet the Contemporary Period is concurrently associated with another shift in focus: to inferences made on how subjects are thinking about or interpret probability.

The Contemporary Period

Research on the CLT from the Post-Piagetian phase and from the transition phase built the foundation for research on the CLT in mathematics education. Building upon research from the previous period, a number of researchers (Ayton, Hunt, & Wright, 1991; Batanero, Green, & Serrano, 1998; Batanero & Serranno, 1999; Cox & Mouw, 1992; Falk & Konold, 1997; Konold, 1989, 1991, 1995; Konold, Pollatsek, Well, Hendrickson, & Lipson, 1991; Konold, Pollatsek, Well, Lohmeier, & Lipson, 1993; Lecoutre, 1992) focused on the creation of cognitive models that were meant to interpret how subjects were thinking about (i.e., interpreting) probability during their completion of tasks such as the CLT. These new models, inferred from researchers interpretations of CLT responses, developed into what Jones and Thornton (2005) called “further research on heuristics” (p. 75). Specific to the CLT in this period were Konold’s (1983, 1992) *outcome approach*, and LeCoutre’s (1992) *equiprobability bias*.

The outcome approach. Research on the CLT during the Contemporary Period was dominated by Konold (Falk & Konold, 1997; Konold, 1989, 1991, 1995; Konold et al., 1991; Konold et al., 1993). In fact, Konold’s research even “extends into the next” (Jones & Thornton, 2005, p. 75) period. Konold (1991) decided, “rather than exploring how people arrive at probabilistic judgments, [he was] interested in how people interpret a question about probability" (p. 146). However, the author contends that Konold's focus was not necessarily on subjects' interpretations of the question, nor on responses or answers of the task. Instead, "a model of informal reasoning under conditions of uncertainty, the *outcome approach*, was developed to account for the nonnormative responses of a subset of 16 undergraduates who were interviewed" (Konold, 1989, p. 59). Consistent with the CLT response interpretations analyzed by researchers throughout the periods discussed, "[Konold's] purpose [was] to demonstrate how statements that otherwise would be regarded as contradictory or incomprehensible can be understood if one assumes that students are reasoning according to [a different interpretation of probability,] the outcome approach" (Konold, 1991, p. 147). Application of this theoretical model to CLT responses demonstrated a new interpretation for how subjects were interpreting probability. In essence, Konold recognized that normative interpretations and heuristic interpretations of probability did not capture the multivalence associated with the CLT, and probability in general.

Konold (1995) went further and claimed that the multiple models—normative, heuristic, and informal—could conflict in responding to a question such as the CLT. Konold et al. (1991) in examining for consistencies over different problems found "switching among alternative perspective[s] of uncertainty" (p. 360). Further complicating the matter, "different perspectives can be employed almost simultaneously in the same situation" (p. 360) because "people use a variety of frameworks and beliefs concerning uncertainty" (p. 361).

To explicate the point, Konold et al. (1993) gave students a most likely version of the CLT followed by a least likely version. It was found that for the most likely version some subjects answered using the outcome approach, and for the least likely version subjects answered using the representativeness heuristic. Having demonstrated individuals' ability to have different problems cue different knowledge Konold et al. concluded "in one problem, a person may appear to reason correctly, but in another, this same person may reason in ways that are at variance with probabilistic and statistical theory" (p. 393).

Further to the issues of interpretation of probability presented, Konold also raised concerns about the inherent flaws associated with the task itself. For example, assessing probabilistic reasoning, "based on [the] correct performance on a few multiple-choice items are not necessarily indicative of a normative understanding" (Konold et al., 1991, p. 362). Also, inherently flawed is "the difficulty of assessing conceptual understanding with multiple-choice items" (Konold, et al., p. 392). Perhaps most poetically put, "there is no simple story about how students reason about chance" (Konold et al., 1993, p. 413). Moreover, and in a recognition of the multivalence of the CLT, "to account for these types of inconsistencies, it is critical to understand the beliefs and reasoning processes that underlie the various answers that subjects give" (Konold et al., 1993, p. 393). While the research of Konold dominated the Contemporary period of probability education, other research was also being conducted.

The equiprobability bias. Lecoutre's (1992) research led to another bias, which should, according to her, "be added to the list" (p. 558) of heuristics and biases from psychology. More specifically, Lecoutre's research was based on interpreting CLT responses that involved a relationship between randomness and equiprobability. Lecoutre (1992) declared, "random events are thought to be equiprobable 'by nature'" (p. 557). For example, two sequences of coin flips, one with three heads and two tails and the other with four heads and one tail, would be considered equally likely because flipping a coin is a random process and thus "the two results to compare are equiprobable because it is a matter of chance" (p. 561). Interpreting responses via the equiprobability bias demonstrates that correct answers to the CLT (i.e., all sequences are equally likely) can be derived from incorrect reasoning (i.e., random events should be equiprobable by nature) of the task.

The Artefactual period

Researchers' incessant reliance on inferences, hypotheses, and theories when interpreting CLT responses—whether made from answers, responses, questions, or cognitive models—was under scrutiny in the transition period between the Contemporary and Artefactual period. For example, Hirsch and O'Donnell (2001) heeded the call put forth by Shaughnessy (1992) and Konold (1991) for research to look further into the reliability of the CLT. In general, the Artefactual period can be characterized as reflectively critical. By reflective it is meant that research at this time "looks back" and analyzes research conducted on the CLT, and by critical it is meant that the reflections critique earlier research.

Two areas of critique that dominate the move into the Artefactual Period are (1) the inferential nature of the forced response items found in the psychology literature, and (2) the most likely versus least likely versions of the CLT upon which Konold developed the outcome approach. Interestingly, the inferential critique of the CLT's forced response items appears to garner support as the years pass. For example, Shaughnessy (2003) stated, "Kahneman and Tversky did not give the choice of [equally likely], nor did they ask subjects to give a reason for their answers in their early studies. They forced their subjects to make a 'most likely' choice among one of these equally likely binomial sequences" (p. 219). While

Shaughnessy specifically calls upon the two items that he later changed, the critique of the CLT also occurs in a general sense: “Kahneman and Tversky’s (1972) representativeness heuristic is so widely cited in the literature as an explanation for how or why people make certain probabilistic determinations that it often goes unchallenged in terms of its explanatory power” (Rubel, 2007, p. 533).

Ironically, the research of Konold—that Liu and Thompson (2007) make reference to in their critique of Tversky and Kahneman’s work—is also undergoing scrutiny. The distinction in results Konold (1989) found for subjects answering the most likely versus the least likely version of the task appears hard to replicate, and efforts to replicate results add inconsistent results. While reflecting on the research of Konold et al. (1993), Tarr and Lannin (2005) state, “clearly a substantial number of students who demonstrated some understanding of independence in the most likely case abandoned this thinking in the least likely case” (p. 220); and while Jones, Langrall, and Mooney (2007) further bolster the results of Konold et al. (1993) by stating that Watson, Collis, and Moritz’s (1997) finding were similar, Rubel (2006)—in not just a reflection, but an attempt to replicate findings earlier produced—notes that there were “very few instances of such inconsistencies” (p. 55) between the least likely and most likely versions of the task.

However, beyond reflection and critique, research on the CLT (Abrahamson, 2008; Hirsch & O’Donnell, 2001; Rubel, 2006, 2007) also continues. In this research, there is (1) a renewed focus on subjects’ reasoning, and (2) new interpretations of responses that attempt to better explain subjects’ reasoning. The research of Rubel (2006, 2007) aligns with the former point, and Abrahamson (2008) aligns with the latter.

Rubel’s (2006, 2007) research provides a nice balance of research and reflection indicative of the Artefactual Period. Her study "focuses specifically on the reasoning, strategies, and underlying cognitive models of middle school and high school students and offers description and analysis of students' mathematically correct and incorrect responses" (2007, p. 536). When subjects were presented with the sequences HTHTHT, HHTHTT, HHHTTT and THTTTT, Rubel (2006) found that while younger students mostly employed the outcome approach, the older students were, for the most part, reasoning correctly. Moreover, the sequence HHTHTT appeared the most representative because of the ratio of heads to tails and the seemingly random order of the sequence. However, and in an unlikely—according to prior research—turn of events, HTHTHT was the most likely sequence to occur for thirteen percent of the sample; unlikely, because according to prior research, a perfect alteration of heads and tails often demoted the sequences likelihood. Further, and as mentioned, there were only a few instances of differences between the least likely and most likely versions of the task, unlike in Konold (1989). In reflecting on the aforementioned findings, Rubel (2007) describes that a number of different methods were used to get to the answer "equally likely." However, reasons for the correct answer fell into different justifications, of which some were right and some were wrong. In general, the most common response for a correct answer of equally likely was begotten from incorrect reasoning.

The multivalence of correctness associated with CLT responses, which, it can be argued, began with Konold's attempt to account for absurd responses to probability questions and then continued with Borovcnik and Bentz's interpretations, has also been addressed in the Artefactual period. Abrahamson (2008), embracing the multivalence of correctness also seen with Borovcnik & Bentz (1991), declares an individual who states HTTH as more likely than

HHHH is not incorrect because “they are counting outcomes yet ignoring their order, thus successfully comparing the likelihoods of the order-less events ‘2H-2T’ and ‘4H’” (p. 10). They are incorrect that HTTH is more likely than HHHH from a normative solution perspective, but if it is the case that they are in fact answering the “order-less” version of the question, they are considered correct in both their reasoning and their answer.

As seen with the literature review presented, research on the CLT is extensive. In fact, research on the task has been taking place for almost thirty years. In recognition of these points the next chapter is dedicated to synthesizing the literature review in order to help explicate the author’s proposed focal shift for research on the CLT.

CHAPTER TWO

An Interpretation of the Literature

The research literature reviewed in the previous chapter presented a variety of interpretations for CLT responses. In this chapter, a CLT Response Interpretation Framework, within which the reviewed literature can be more coherently situated, is developed and used by the author as a tool to help interpret CLT response interpretations. However, and to begin the chapter, CLT theories or models created to account for subjects' responses in past and present literature is explored.

CLT Theories

Shaughnessy (1992) declares that psychologists "customarily…attempt to explain what they observe on the basis of some theoretical model" (p. 469). The present author further notes that the models in CLT research (whether found in psychology or mathematics education) are based on *abductive reasoning* (Peirce, 1931). Upon hearing or reading a response Q given by a participant, a researcher infers a rule P which implies Q (e.g., if the student is using the outcome approach, then she will answer Q) and a consequence P (e.g., that the student is using the outcome approach). Peirce describes abductive inferences as inferences to the *best* explanation.

Particular hypotheses that garner support to be considered true subsequently reveal themselves as theories or models found in the mathematics education and psychology literature. However, and despite the overwhelming evidence for certain models (e.g., outcome approach, equiprobability bias, representativeness) developed to help explain individuals' reasoning when dealing with questions under conditions of uncertainty, one cannot declare for certain that an individual is in fact answering the CLT with, for example, the representativeness heuristic. Nevertheless, the models have gained enough support that they saturate CLT research literature. Moreover, the saturation has occurred to such a degree that often the models are misconstrued as matter-of-fact statements such as, "the student was using the outcome approach." In other words, the initial abductive inference has turned the rule hypothesized in earlier research into a statement of truth accepted by subsequent research.

Consider, for the moment, the following situation: If one studies hard, then one gets good grades. Just because one gets good grades does not necessarily mean that one studied hard. For example, one may have gotten good grades because one cheated. Thus, it is more appropriate to declare that if one gets good grades the most probable explanation is that one studied hard, which may or may not be the case. Similar situations are littered throughout the CLT literature. For example, if one uses the equiprobability bias when answering the CLT, then one may choose all sequences equally likely. However, because one chose all sequences equally likely does not necessarily mean that one used the equiprobability bias. One may have chosen all sequences equally likely based on normatively correct reasoning.

The observation of good grades does mean that one cannot declare with certainty that studying hard is the rule used for explanation. However, through the process of abductive reasoning one can proceed to hypothesize that studying hard was the reason for the good grades, and if the hypothesis were true, then the achievement of good grades would follow suit. Consequently, there would exist reason to suspect that studying hard, the hypothesis, is true, which could be denoted the study-hard theory.

In a more general sense, facts are used as a starting point, a particular hypothesis—derived from inferences and used to best explain the facts observed—is presented, and if it is the case that if the hypothesis were true it would best or most likely explain the observed facts, there exists reason to suspect the theory hypothesized is true (Lipton, 1991). CLT theories are created in a similar fashion. For example, Lecoutre (1992) noticed that individuals declared all sequences to be equally likely to occur because flipping a fair coin is random and anything can happen. This observed fact acted as the starting point. Then it was hypothesized that certain individuals answering the question employed the equiprobability bias. If it were the case that certain individuals were answering via the equiprobability bias then it would follow suit that those individuals would in fact answer the CLT equally likely because flipping a fair coin is random. As such, there existed reason to suspect that the equiprobability bias is true. Further, the equiprobability bias has subsequently become a theory or model to explain how certain individuals respond to probability tasks. As more and more evidence mounts for models such as the equiprobability bias, there appears to be—on behalf of the research community—more and more reason to suspect that the equiprobability bias is true. To reiterate, it cannot be said for certain that if an individual answers that all sequences are equally likely to occur that they have done so according to the equiprobability bias. In fact, it cannot be claimed with certainty that individuals, any individuals, employ any of the cognitive models known when answering the CLT. Hypotheses like the representativeness heuristic and the outcome approach can be seen as (1) models or theories hypothesized to explain observed results, and (2) as "new research" created through the process of abduction resulting from analyzing CLT responses. A similar approach will be taken with the subjective-sample-space, but certain frameworks must be presented first.

Response Interpretation Framework (RIF)

Consider, for the moment, the hypothetical response to the CLT shown in Figure 3, which consists of two components: *multiple-choice selection* and *justification* (i.e., the written justification for the multiple-choice answer selected).

> *I chose e) all sequences are equally likely to occur. Flipping a coin is a random process, which means that anything can happen, so all of the sequences are equally likely to occur.*

Figure 3. Hypothetical CLT response

The choice e) evidences which multiple-choice answer has been selected, and the statement concerning coin flipping and random process evidences the participant's justification for their multiple-choice selection, or answer. Despite the clarity associated with the CLT response components (i.e., answer and justification), CLT responses—like the task itself—are multivalent. Moreover, the different interpretations are derived from the answer component, the justification component, and the union of the two components.

Multivalent elements of CLT responses

The different components to the CLT responses dictate the multivalence be examined in terms of different elements. Specific to this research, the multivalence of four elements of the Response Interpretation Framework are examined: (1) probabilistic perception element, (2) the answer element, which corresponds to the multiple-choice selection component of the task, (3) the answering element, which corresponds to multiple-choice selection justifications, and (4) through the combination of the answer and answering elements, the interpretations of correctness/incorrectness associated with CLT responses.

(1) Probabilistic perception element. The answer element of the CLT is, by acclamation, considered correct if the normatively correct answer—all sequences are equally likely to occur—is chosen. However, what is meant by normative is a matter of interpretation. While "'normative' is used…to refer to some theoretical model for assigning probabilities or likelihoods to events" (Shaughnessy, 1992, p. 470), there are different theoretical models for assigning probabilities to chose from. In fact, there are (at least) four different theoretical models for assigning probabilities: classical, frequentist, subjective, and "formal" (Hawkins & Kapadia, 1984, p. 349). In terms of the theoretical model for assigning probabilities, and concurrently the normatively correct solution for the CLT, researchers in psychology—interested in the tension between axiomatic probability and human judgment—and mathematics education—reflecting the curricular concentration of one interpretation of probability—have adopted the classical interpretation of probability. Thus, it is perhaps more appropriate to state: Based on the intended interpretation of researchers (i.e., adoption of the classical interpretation of probability for the CLT), the normatively correct answer to the CLT is that all four sequences are equally likely to occur. Nevertheless, multivalence remains for CLT response interpretations, and the multivalence of the answer element is compounded by different interpretations of probability that may be employed in CLT responses

(2) Answer element of CLT responses. Interpreting responses to the CLT from a normative solution (i.e., answer element) perspective, leads to one of two possible conclusions, as shown in Figure 4.

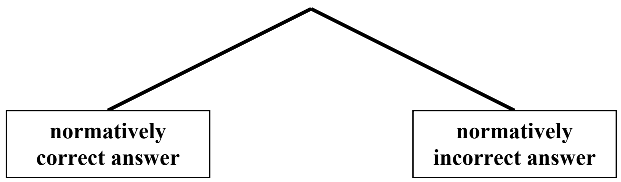

Figure 4. Answer element distinction for CLT responses

First, an individual who declares that all sequences are equally likely to occur has answered the CLT correctly. Second, an individual who declares one particular sequence to be less likely (or more likely) than any of the others presented is incorrect. However, correctness is supposing that individuals correctly arrive at the correct answer and incorrectly arrive at the incorrect answer. For example, an individual who correctly arrives at the correct answer, and a second individual who guesses—presented here as an incorrect approach for the sake of argument—the correct answer are one in the same. Similarly, an individual who incorrectly arrives at the incorrect answer, and a second individual who correctly arrives at the answer, yet makes a "selection error" in the final stages of completing the task, or is correct in certain parts of their answering of the task but eventually arrives at an incorrect answer, are one in the same. As such, the answer element of responding to the CLT is multivalent. Moreover, interpreting an individual's response with the normative solution can only determine whether or not the individual's answer is correct or incorrect and provides no information about how the question was answered.

Interpreting a CLT response with only the answer element leads to further multivalence for the answer element. For example, in initial research on the CLT (e.g., Tversky & Kahneman, 1974), justification of responses was not included as part of the task. Thus, correct responses to the CLT (i.e., declaring that all sequences are equally likely to occur) coexisted as (1) the normative solution, and (2) a gauge to determine whether or not an individual had achieved the researchers' intended interpretation of the task. However, it must be noted that there are inherent issues with accepting the normative solution as the determinant of whether or not the intended interpretation has been achieved. It is incorrect to assume (1) that individuals who have correctly answered the CLT have achieved the intended interpretation of the researcher, and (2) individuals who have incorrectly responded to the CLT have not achieved the intended interpretation of the researcher; because (1) if an individual has correctly answered the CLT that does not *necessarily* mean that they have achieved the intended interpretation, and (2) if an individual has incorrectly answered the CLT that does not *necessarily* mean that they have not achieved the intended interpretation and would wind up with different probabilities.

(3) Answering element of CLT response. The reasoning or approach—variably represented in probability education literature as: "reasonable mathematical thinking" (Rubel, 2006, p.50); "model" (Lecoutre, 1992, p. 557); reduction of degrees of freedom (Borovcnik & Bentz, 1991); event based human intuition (Abrahamson, 2008); formal versus natural

cognition (Konold, 1989); strategies (Green, 1983)—associated with answering probability tasks is not addressed in the answer element of CLT responses.

Given the multitude of names associated with arriving at the answer, the author contends that (1) *answering* be adopted to encapsulate, in a liberal sense, the process of arriving at an answer, and (2) the answering element of CLT responses is multivalent. The author further wishes to incorporate the answering element as part of the response interpretation for the CLT, and that the answering element be broken into two categories: *correct answering* and *incorrect answering*, as seen in Figure 5.

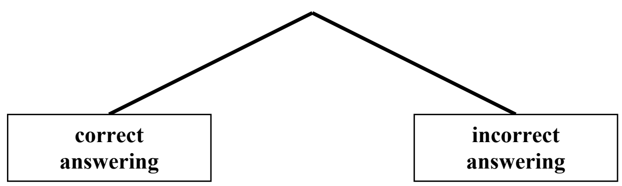

Figure 5. Answering distinction for CLT responses

(4) Correctness/incorrectness element. Allowing response interpretations to include both the answer and answering elements of responses to the CLT allows for investigation into two further outcomes once veiled by the multivalent answer element: (1) a correct answer to the CLT begotten under false pretenses, and (2) an incorrect answer to the CLT, which possesses correct elements of answering the CLT. Recognition of these new interpretations will provide insight for examination beyond the assumption that correct answers are based on correct reasoning and incorrect answers are based on incorrect reasoning. Moreover, the emergence of two new outcomes coupled with the two original outcomes implies that there are (at least) four branches (shown in Figure 6) from which responses to the CLT can be interpreted.

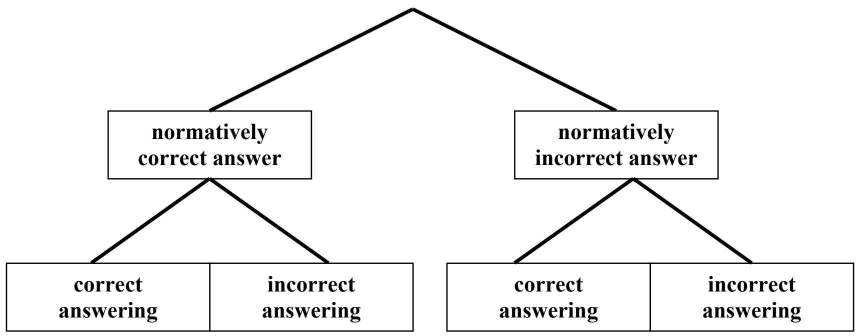

Figure 6. Answer and answering elements distinction for the CLT

First, an individual's correct answer is based on correct answering of the task (i.e., correct answering has led to a normatively correct answer). Second, an individual's correct answer is based upon incorrect answering of the CLT (i.e., an individual uses contextually inappropriate answering, yet, somehow, arrives at the normatively correct response). Third, an incorrect answer is based upon incorrect answering of the CLT (i.e., incorrect answering on the individual's behalf manifests itself in a normatively incorrect response). Fourth, an individual who answered normatively incorrectly may have correctly answered elements of the problem, but not the problem itself (i.e., while an individual's approach to answering the task possesses elements of correct answering, their final answer is normatively incorrect).

Incorporating the answer/answering element for responses to the CLT does not reduce the multivalent nature of response interpretation. In fact, by allowing for response interpretations to include answer and answering elements of the CLT, the notions of correctness and incorrectness become a matter of interpretation. To exemplify the aforementioned assertion, the "branches" of correct answers begotten from incorrect answering, and incorrect answers begotten from correct answering are examined.

The author asserts that correct answers to the CLT based upon incorrect answering of the task are both correct and incorrect, thus they are multivalent. Correct in the sense that the question has been answered correctly, and incorrect in the sense that the answer was gotten under alternative motivations. Thus, the notion of correctness or incorrectness is a matter of interpretation. An individual whose focus lies solely with the answer to the CLT interprets the answer as correct. However, a second individual whose focus lies solely with the answering of the CLT may interpret the answer as "correct with an asterix," or may interpret the answer (which includes the answering) as incorrect. Treating only the left branch of Figure 6 (correct answer and correct answering) as correct oversimplifies the problem for those interested in human thinking.

Similarly, incorrect answers to the CLT based upon (partial) correct answering of the task are also both correct and incorrect, thus they too are multivalent. For example, an individual is incorrect in stating, for five flips of a coin, that THTTT is less likely that HHHTT, because, as explained, each of the sequences is equally likely to occur. That said, the author wishes to contend that if the individual were to have interpreted the answer element of the task in terms of the number of heads and tails, that particular individual would be correct in that obtaining four tails and one head is less likely than obtaining three heads and two tails for five flips of a coin, even though they present their notion through declaring THTTT is less likely than HHHTT. The caveat is that the individual has not answered the intended interpretation of the CLT and, as such, is incorrect. However, in recognizing that they have interpreted and subsequently answered a different question than the one intended, they are correct in their answer and their answering of their interpretation of the question. Thus, whether the answer is correct or incorrect is, again, a matter of interpretation.

While the correct answer to the CLT is that "all sequences are equally likely to occur," research is beginning to suggest that individuals who incorrectly answer the CLT may, in fact, not be incorrect. For example, Abrahamson (2008) states, "people who judge a HTTH 4-coin flip as more likely than the equiprobable HHHH flip are not in error" (p. 8). He continues on to mention, "they are counting outcomes yet ignoring their order, thus successfully comparing the likelihoods of the order-less events '2H—2T' and '4H'"(p. 8). In other words, and stated in terms of the framework elements presented in this chapter: An individual who declares HHHH as less likely than HTTH is incorrect if interpreted from an answer element perspective, but if interpreted from an answering element perspective they are correct, in that the flipping four heads is less likely than the event flipping two heads. In fact, flipping four heads (HHHH) is one fourth as likely as flipping three heads and one tail when order is not taken into consideration or when flipping four coins at once (HHHT, HHTH, HTHH, THHH).

CLT response multivalence

As shown, the CLT response is multivalent; however, and as also shown, the multivalence of the response is, or can be, based on a multitude of different elements presented. In other words, the multivalence of the response element can be derived from the answer, probabilistic perception, answering, correctness/incorrectness elements of the task, whether treated in some manner of inter-elemental unison or singularly. For example, a participant who chooses one particular sequence to be less likely because they employ a frequentist or experimental interpretation in their justification may be considered both correct and incorrect depending on one's interpretation of the response.

CLT response in superposition

Given responses to the CLT are multivalent, the response interpretations may be described as existing in a state of superposition, where no one interpretation of the response exists because all response interpretations concurrently exist. In order to collapse the CLT response from a state of superposition to a particular interpretation requires an observation. In other words, one researcher may interpret the CLT response in terms of the answer chosen (e.g., Tversky & Kahneman, 1974), while another individual may interpret the response in terms of the answer and justification (e.g., Shaughnessy, 1977), while yet a third individual may interpret the response solely in terms of the justification. To determine which CLT response interpretation collapse has occurred, inferences can be made through examination of research literature developed by individuals who have interpreted CLT responses. In doing so, examination of the interpretation of CLT responses will not only provide insight into the different interpretation collapses for different individuals, but will also show how factors such as academic field, particular intended interpretations of the CLT and judgment thereof, inferences made on comments, and implicit assumptions all play a role in the hypotheses that have developed from interpreting CLT responses. As was the case with the literature review, the interpretation of the literature begins with the field of psychology, and then turns to mathematics education.

Interpreting CLT response interpretations

As presented, CLT response interpretations may be described as existing in a state of superposition, but the author's review of the literature, seen in the first chapter, has, in essence, collapsed the state of superposition and shown that different individuals interpret CLT responses in different ways. Moreover, the observation of the literature has shown how the hypotheses associated with different interpretations correspond to different cognitive models (e.g., representativeness, the outcome approach).

Also shown, the use of inferences on CLT responses attempt to clarify how the CLT was interpreted. At present, and in an analogous fashion, inferences will now be made on the CLT literature in an attempt to clarify how different researchers interpreted CLT responses and consequently obtain an alternate picture of the CLT research landscape as presented in Chapter One.

Interpreting psychological CLT response interpretations

As witnessed in the review of the literature, analyses of the answer component of the CLT by psychologists, at the time, led to one of two results. First, individuals who answered the CLT correctly were deemed to have approached their answer to the question correctly (perhaps more appropriately stated: individuals who answered the CLT correctly were not the focus of investigation, especially when building a body of work on heuristics and biases). As such, whether or not correct answers were begotten from correct or incorrect answering was a moot point for researchers. Second, it was determined that individuals who answered the CLT incorrectly were considered to be employing a heuristic approach or model, the representativeness heuristic: considered as an incorrect but understandable answering of the CLT, i.e., incorrect answers to the CLT were seen as incorrect as a result of incorrect answering with the representativeness heuristic. However, while it was determined *that* individuals were answering the task incorrectly, it was inferred—from the forced response items devoid of reasons for choice of sequence—why they were answering the task incorrectly. Considering that the inferences made may have been incorrect, it is more appropriate to declare that individuals answered the CLT incorrectly, which may stem from use of the representativeness heuristic, instead of declaring that the representativeness heuristic is *why* individuals incorrectly answered the CLT.

In a general sense, psychological research focused on (1) correct answers with the assumption of correct answering, and (2) incorrect answers with the assumption of incorrect (i.e., heuristic) answering, as seen in Figure 7. Alternatively stated, the focus of research for psychologists can be described as using but not recognizing the multivalence of the answer element of the CLT responses.

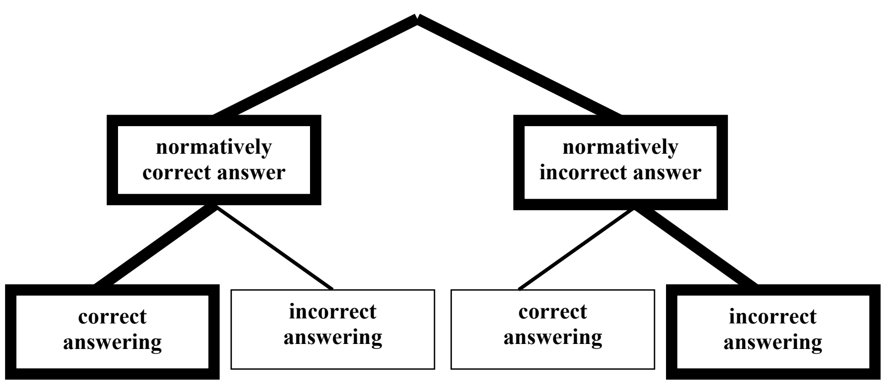

Figure 7. Organization of psychological CLT response interpretations

Interpreting mathematics education CLT response interpretations

While psychologists did not explicitly recognize the multivalence of the answer element of CLT responses, researchers in mathematics education embraced the multivalence found in the: answer (Shaughnessy, 1977, 1981), answering (Rubel, 2006, 2007), probabilistic perception (Konold, 1989, 1991; Lecoutre, 1992), correctness/incorrectness (Borovcnik & Bentz, 1991; Abrahamson, 2008) elements of CLT responses. Consequently, researchers in mathematics education have investigated all the multivalent elements of CLT responses. Moreover, different interpretations of responses were based on which multivalent element of the response was under investigation. For example, Konold's (1989) outcome approach resulted from inferences made on the multivalence of the probabilistic perception employed by individuals responding to the CLT.

Given mathematics education researchers interpreted responses through the multivalent elements of CLT responses, mathematics education research can be described as focused on (1) correct answers derived from incorrect answering, and (2) incorrect answers derived from correct answering, as seen in Figure 8.

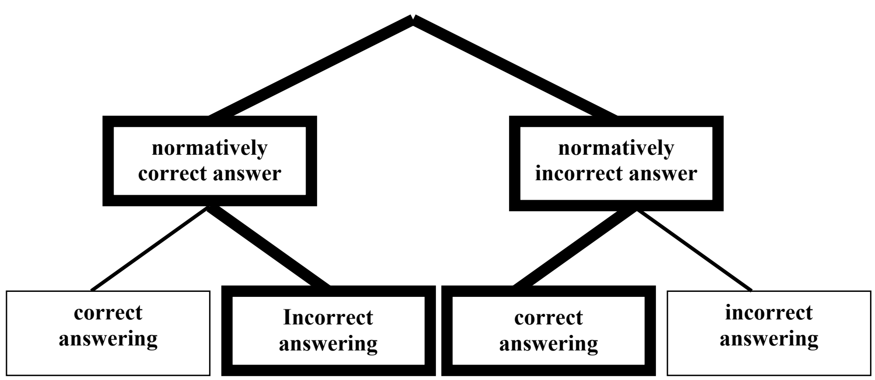

Figure 8. Organization of mathematics education CLT response interpretations

The "branch" of correct answers yet incorrect answering has a focus on the answering element (e.g., Rubel 2006, 2007), and is associated more with the multivalence of the answering element of CLT responses. Alternatively, the branch of incorrect answers that possess correct answering is seen in the work of Borovcnik and Bentz (1991), Konold (1989, 1991), Lecoutre (1992), Abrahamson (2008), and is associated more with the multivalence probabilistic perception element of CLT responses. Further, both branches, when also incorporating the answer element of CLT responses into the research, are associated with the notions of correctness/incorrectness as seen in the works of Borovcnik and Bentz (1991) and Abrahamson (2008).

Interpreting psychological and mathematics education CLT response interpretations

By combining the research found in psychology, which did not focus on the multivalence of the CLT responses, and mathematics education, which did focus on the multivalence of the CLT responses, each of the four "branches" found in Figure 8 have, at some point, been the focus of investigation. More specifically, psychology research utilized the answer element of CLT responses, while mathematics education utilized the probabilistic perception, answering, and correctness elements of CLT responses. However, and despite the amount of research that has been conducted on the CLT, the rampant multivalence dictates that interpretations for responses to the CLT, other than those found in the literature, may be possible; and examining an alternative interpretation of CLT responses will be the focus of the next chapter.

CHAPTER THREE

Study I

In the previous chapter the author's Response Interpretation Framework demonstrated that despite the wide variety of research presented on the CLT an overarching theme emerges: prior research on the CLT consists of interpretations and hypotheses based on multivalent elements of CLT responses. In a general sense, elements of the CLT *response* are interpreted by *researchers*, whereas elements of the *task* are interpreted by *respondents*. However, it is possible for researchers to attempt to interpret responses for evidence of how respondents interpret the task. Motivated by the notion that respondents may have interpreted that task differently than intended and interested in the interpretation of the task by respondents, the author proposes a focal shift: instead of continuing to base research on multivalent elements of *CLT responses*, the author creates a new theoretical framework, the CLT Task Interpretation Framework, with which to analyse CLT responses based on multivalent elements of the *CLT*. After an explication of the novel theoretical framework, the results and analysis of Study I are presented and followed by a conclusion and discussion.

Task Interpretation Framework (TIF)

The CLT response interpretation framework has shown exploring the multivalence of the *CLT response elements* has been extensive in CLT research, yet exploring the multivalence of the *CLT elements* has not followed suit. Consequently, a new approach is not to examine the multivalence of *CLT response elements*, but to examine the multivalence of *CLT* *elements*. The author hypothesizes that CLT responses can be used to interpret how respondents are interpreting and answering the CLT. The author further hypothesizes the focal shift, from response interpretations to task interpretations, on the part of the researcher, will provide new insight into probabilities associated with perceptions of randomness. However, before the hypothesis can be tested, a new theoretical framework for analysis is presented.

Multivalent elements of the CLT

The CLT is multivalent. Specific to this research, four elements of the task—sequence, likelihood, experiment, and question—will be investigated, and will be described below in accordance with an example of the CLT seen in Figure 9.

> *Which of the following sequences is the least likely to occur from flipping a fair coin five times. Justify your response.*
> *a) THTTT*
> *b) THHTH*
> *c) HHHTT*
> *d) HTHTH*
> *e) all four sequences are equally likely to occur*

Figure 9. CLT example for elemental examination

First, potential answers to the task—the four sequences of heads and tails derived from flipping a fair coin (i.e., THTTT, THHTH, HHHTT, HTHTH)—will be investigated. Second, analysis turns to the use of the word likelihood found in the sentence components of the task—the question posed in the task (i.e., "which of the following sequences is the least likely to occur from flipping a fair coin five times") and the equally likely response (i.e., "all four sequences are equally likely to occur"). Third, attention will be paid to possible interpretations of the probabilistic experiment of flipping a fair coin five times. Fourth, different interpretations of the question will be considered.

(1) Sequence element. Consider the four sequences presented in the CLT: While each of the sequences can, and is intended to, be seen as a sequence of heads and tails for five flips of a coin (i.e., THTTT interpreted as: tail on the first flip, then head on the second flip, and then tail on the third, fourth and fifth flips), the four sequences may be interpreted in other ways. For example, THTTT, THHTH, HHHTT, and HTHTH are, concurrently with the first interpretation, sequences with a ratio of four tails to one head, three heads to two tails, three heads to two tails, and three heads to two tails, respectively. From this interpretation, the sequence THTTT has a different ratio of heads to tails (4:1) than the sequences THHTH, HHHTT, and HTHTH (3:2). As such, THTTT is different than the others. The ratio of heads to tails is not the only plausible interpretation of the sequences presented in the CLT. There are other attributes associated with sequences of heads and tails derived from flipping a fair coin that may also be concurrently employed when interpreting sequences.

Specific to this research, the number of switches (or alterations), the length of the longest run, and the combination of the two attributes (i.e., switches & longest run) are examined. As such, the sequences of THTTT, THHTH, HHHTT, and HTHTH, which coexist with the ratio of heads to tails interpretation, also coexist as: (1) sequences that have two switches, three switches, one switch, and four switches, respectively; (2) sequences that have a longest run of three, two, three, and one, respectively; and (3) sequences that have two switches and a longest run of three, three switches and longest run of two, one switch and a longest run of three, and four switches with a longest run of one, respectively. As evidenced, there exist a variety of meanings for the sequences of H's and T's that are seen in the sequence element of the CLT. The sequence element of the CLT is multivalent.

(2) Likelihood element. Similar to the sequence element of the task, the likelihood element of the task is multivalent. While a number of words found in the sentence component of the task (e.g., sequence, occur) are possible sources of investigation, it is not necessary to investigate all of the words to evidence the likelihood element of the task is multivalent. As such, and at the author's discretion, investigation will focus solely on the notion of *likelihood*, more specifically, the colloquial use of the word likelihood. For example, 'in all likelihood Jim has passed his exam.' In this particular use of the word 'in all likelihood' implies that it is probable, or perhaps very probable, that Jim will have passed his exam. That being said, 'very probable' is yet another colloquial usage of the notion of probability, which can be closely connected to 'certainty.' This procedure, of changing from one colloquial word for probability or likelihood to another adds to the multivalence of the task. Further, there is tremendous usage of colloquial terms for probability in the English lexicon, and usage of these words begins at a very young age also adding to the multivalence of the task. As such, the multivalence of the colloquial representations of likelihood may bring high levels of coexistence of interpretations to the likelihood component of the CLT.

(3) Experiment element. The experiment element for the CLT is the flipping of a fair coin five times, and as will be shown, contains a number of coexisting interpretations. While the experiment may not appear as a source of much confusion, there exist a number of plausible interpretations. More specifically, the derivation of the four particular sequences is one

manner in which it can be shown that the experiment element of the CLT is multivalent, because interpretations of how the sequences were derived for the task are not entirely clear. The sequences could have been derived from: flipping one coin twenty times, or twenty coins all at once. There are other plausible combinations of coin flips and coins: five coins flipped at once on four separate occasions, four sets of five different coins flipped all at once, four different coins flipped one at a time for five times simultaneously, four different coins flipped subsequently one at a time for five times, or four sets of five coins flipped simultaneously or subsequently one at a time. Given coin flips are independent events it would not matter how the outcomes were derived; however, it must be recognized that independence is one of the key elements being examined in the implementation of the CLT. As such, it is plausible that an individual answering the task may be influenced by the derivation of sequences.

As shown, the sequence element of the task, the likelihood element of the task, and the experiment element of the task are multivalent. However, the argumentation presented above is all based on the assumption that the question posed in the task is being appropriately interpreted. As such, the question element of the CLT is now investigated.

(4) Question element. There are viable, coexisting interpretations of the question posed in the CLT. For example, the question does ask the subject to determine which of the sequences presented is least likely to occur, but sequences presented may be pitted against the other three sequences in the task for likelihood comparison, or may be pitted against all thirty-two possible outcomes for five flips of a fair coin. That said, the coexisting interpretations do not have to be based solely on the question element of the task.

The multivalence of the question element of the task can also be derived from the answer, likelihood, and experiment elements of the task, whether treated in some sort of unison or singularly. Alternatively stated, and as an example, an unintended interpretation of the sequence element of the task implies an unintended interpretation of the question element of the task. If a subject interprets the sequences of heads and tails as a ratio interpretation, that individual is answering an unintended interpretation of the question element of the task. As another example, if an individual interprets likelihood in an unintended manner, then they are also not interpreting the question element of the task in an intended manner.

The task of achieving the intended interpretation for the question element is based on achieving intended interpretations for the answer, likelihood, experiment, and question elements, treated singularly or in some type of inter-elemental arrangement. While examination of the CLT's multivalence occurred for different elements of the task treated separately, inter-elemental examination could have also been conducted. However, it is not necessary, and could be argued futile to attempt to further the point that the CLT is multivalent. Instead, investigation turns to the ramifications of the CLT being multivalent. Given the argumentation for the multivalence of the elements of the CLT, the author contends that the question element of the CLT is perilously multivalent. Probabilistically stated, the chances of answering the CLT as the researchers intended is unlikely. Alternatively stated, it is likely that a subject answering the CLT is answering an unintended interpretation of the task, and the results of Study I and Study II will provide evidence in support of this assertion.

CLT in superposition

Being multivalent, and as shown, the CLT has a number of possible (coexisting) interpretations. Or, in other words, the CLT—like the statement “the shooting of the hunters was dreadful” and the CLT responses—can be described as existing in a state of superposition. When in a state of superposition the CLT represents all possible interpretations of the task. As noted before, when in a state of superposition there is no one particular interpretation of the statement, because all interpretations exist at once. In order to collapse the CLT’s state of superposition an interpretation is required. Alternatively stated, to limit the statement to a particular possibility requires an individual to interpret, or complete the task. For example, one individual may, in their completion of the task, interpret the sequence element in terms of the ratio of heads to tails, whereas another individual may interpret the sequence element in terms of the number of switches that have occurred (regardless of whether all other elements are being intentionally interpreted or not).

To determine which interpretation collapse has occurred for an individual completing the CLT, inferences can be made by examining responses made by the individuals who have completed the task. Consider an individual who after completing the task comments, ‘One of the four sequences had a different number of heads to tails.’ The author contends that it is more likely that the individual has interpreted the sequence element of the task in terms of a ratio interpretation. Further, and as another example, the reading of an individual who comments ‘the longer the run of tails the less likely the sequence’ in their justification to the CLT causes the author to infer that it is more likely the individual interpreted the sequence element of the task in terms of the length of runs.

Examination of CLT responses not only provides insight into the interpretation collapse for an individual's completion of the task, but also provides the opportunity to determine whether or not an individual's interpretation matches the intended interpretation of the researcher implementing the task, i.e., achieving the intended collapse. The researcher, knowing the intended interpretation of the task, is able to determine—by the reading of responses made by subjects who have completed the task—whether or not the subject's interpretation aligns with the intended interpretation of the researcher. Ironically, the intended interpretation of the CLT for certain researchers—whether from psychology or mathematics education—has been univalent (i.e., an absence of ambiguity with the intended interpretation).

Study I

Given (1) the potential for new insight into subjective probabilities associated with individuals' perceptions of randomness via the application of the new Task Interpretation Framework, and (2) "although the previously mentioned studies have built a steady trail of evidence that students make errors when evaluating or comparing probabilities of compound events, there remains a need for further exploration of school students' strategies, reasoning, and underlying cognitive models" (Rubel, 2007, p. 533), Study I was conducted to demonstrate that certain individuals interpret the CLT differently than intended by the researcher.

Participants

Participants in Study I were 56 prospective elementary teachers enrolled in a methods for teaching elementary mathematics course, which is a course in the teacher certification program. The 56 participants consist of members from two different classes who were taught by two different instructors. In both classes the variation of the CLT was presented prior to the introduction of probability content in the course.

Task Construction

Inspired by the recurrent development yet consistent findings of the sequence task, Study I provides a new iteration of the comparative likelihood task. Aligned with the task developments presented in the review of the literature in Chapter One, participants were given "all five sequences are equally likely to occur" as an option and participants were also asked to provide reasoning for their response. However, taken into further consideration for this new iteration of the CLT was the ratio of heads to tails in the sequences presented. In four of the five sequences presented (HHTTH, THHHT, HTHTH, THHTH) the ratio of heads to tails was three heads to two tails (3:2). In the second of the sequences presented (HHHHT) the ratio of heads to tails was different at a ratio of four heads to one tail (4:1).

Task

As a result of the information presented above, participants were presented with the following iteration of the CLT seen in Figure 10.

Which of the following sequences is the least likely to occur from flipping a fair coin five times. Justify your response:

a) HHTTH
b) HHHHT
c) THHHT
d) HTHTH
e) THHTH
f) all five sequences are equally likely to occur

Figure 10 CLT utilized in Study I

Results

While there were six choices available to the participants of Study I, responses fell into only four of the six choices. 27 of the 56 participants (approximately 48%) "correctly" chose that all sequences were equally likely to occur, and 29 of the 56 participants (approximately 52%) "incorrectly" chose that HHHHT was the least likely sequence to occur. Further, nine of the participants "incorrectly" chose the sequence HTHTH least likely to occur, and one individual "incorrectly" chose the sequence THHHT as least likely to occur. No participants chose sequences HHTTH or THHTH.

Analysis of results via the Task Interpretation Framework

A preliminary analysis of the sample response justifications in Study I employed the Response Interpretation Framework developed from the interpretation of the literature. Results indicate a number of findings in prior research, as will be demonstrated in certain response verifications. However, in accordance with the scope of the present research, a second analysis of the data, presented in detail below, indicates the multivalence of the likelihood, question, experiment, and sequence elements. Pseudonyms have been used for all individuals presented in the results of Study I (and, for that matter, Study II).

Likelihood element multivalence

As seen from the two selected response justifications presented, and demonstrated theoretically in the author's interpretation of the literature in Chapter Two, while approximately 48% of the participants correctly responded that all sequences were equally likely to occur, not all 48% of participants correctly obtained the correct answer. In fact, and from a further analysis of response justifications, the number of participants who correctly obtained the correct answer was approximately equal to those who incorrectly obtained the correct answer. As such, it can be concluded that, in fact, less than 48% of participants correctly answered the CLT.

Sample response justifications for multivalence of the likelihood element.

Barney: All five sequences are equally likely to occur because when you flip a coin it is random so you cannot predict whether it will turn heads or tails so all these sequences have an equal chance of occurring.

Catie: F) because it is RANDOM!!!!!

As exemplified in the responses of Barney and Catie presented above, Lecoutre's (1992) equiprobability bias was found as one of the incorrect approaches, which led to a correct answer. Lecoutre (1992) declared, "random events are thought to be equiprobable 'by nature'" (p. 557), whereas Catie succinctly described "F) [all sequences are equally likely to occur] because it is RANDOM!!!!!" While correct answers begotten from incorrect answering is a potential area of investigation, which would also further the research into the validity of the CLT (Hirsch & O'Donnell, 2001), the path is beyond the intended scope of the present research. Similarly, further investigation via the Response Interpretation Framework is also a potential area of investigation, but is also not aligned with the author's focal shift (from the multivalence of CLT responses to the multivalence of the CLT). As such, and for the remainder of the analysis of results from Study I, the Task Interpretation Framework will be used exclusively.

Revisiting Catie and Barney's responses via the Task Interpretation Framework demonstrates an interesting relationship between randomness and likelihood: their responses indicate the likelihood element of the CLT is multivalent. As found in Lecoutre's (1992) research random events are considered equiprobable, yet here it is demonstrated that sequences are equiprobable because the process is random. As such, claims made regarding probability could be substituted for randomness and, similarly, claims for randomness could be substituted for probability, without contestation. As a result, there exists a concurrency of interpretations with the term likelihood, thus the likelihood element.

Question element multivalence

Not necessarily evident from an examination of purely numerical results, the multivalence of the question element appears in many of the sample responses justifications presented. Further, different responses are aligned with different multivalent elements of the CLT, as earlier hypothesized. In fact, and according to the inter-elemental argument posed in the Task Interpretation Framework presentation, each of the responses evidencing multivalence in the experiment, likelihood, and sequence elements are also evidencing the multivalence of the question element. Nevertheless, explicit examples of the multivalence of the question element were indicated.

Sample response justification for multivalence of the question element.

Tawnie: Letter T is the least likely to occur for A, B, C, and D. If you count them, T is the least. Because in A there's 5 letters, but T only has two or one so it is the least likely to occur.

Tawnie's response shows that the intended interpretation of the CLT was not achieved and, thus, it can be argued that the intended question was not answered. That said, the intended interpretation of the CLT was not achieved in other responses, yet the evidence of such is more nuanced because it plays on the other multivalent elements of the task, not the question element as seen with Tawnie's response. Nevertheless, Tawnie's response that the letter T is least likely to occur indicates the question element of the task is multivalent.

Experiment element multivalence

Sample response justification for multivalence of the experiment element.

Monah: …it all depends on the coin thrower. Coin throwing is really random

Oliver: You have to think of all the possible things that could happen like the wind could change how much it flips. How it lands also depends on it if it bounces.

Ronald: I think it really depends on how you flip the coin, or where it lands

Nina: because coins rarely flip h, then t, then h repeatedly

Paloma: I was flipping a coin earlier, and it <u>always</u> landed on tails

Quentin: I flipped a coin lots of times and I never got 4 heads in a row.

Monah, Oliver, and Ronald all allude to the experiment element of the CLT. The coin thrower, how the coin is flipped, and certain physical factors such as wind and where the coin lands, are taken into consideration in their answering the CLT. Given the classical interpretation is *a priori* probability, the intended interpretation of the task does not mean for any of the physical characteristics of the experiment to be take into consideration. As such, it can be argued that they too are not answering the intended interpretation of the task. In fact, given the intended interpretation of the CLT uses the classical interpretation of probability, the responses of Nina, which is indicative of a propensity interpretation of probability (i.e., probability is a physical propensity), does not meet the intended interpretation; and, as such, it can be argued that the subjects are not answering the intended question. Similarly, Paloma and Quentin adopt a different theoretical interpretation of probability—the frequentist perspective—than the intended classical interpretation; and, as such, are also not answering the intended interpretation of the task. The unintended interpretations presented all indicate that the experiment element of the CLT interpretation framework is multivalent.

Sequence element multivalence

From the analysis of results the majority of 'incorrect' responses in Study I were based on the multivalence of the sequence element. However, two particular interpretations of the sequence element—the ratio of heads to tails and the perceived randomness, determined by pattern or lack thereof—dominated the response justifications. As such, the response justifications for multivalence of the sequence element are further categorized into (1) the ratio interpretation, and (2) the perceived randomness interpretation.

Sequence element multivalence: Ratio interpretation

Sample response justification for multivalence of the sequence element: Ratio.

Francine: It's most unlikely to have four heads and one tail because there is a 50% chance.

Gerard: It has the "most uneven" amount of heads and tails.

Hannah: B, it's the only one with one T and four Hs, the rest have 3 Hs and 2 Ts.

Given the review and interpretation of the literature, it is not surprising that approximately 52% of the participants incorrectly answered that the sequence HHHHT was least likely to occur. That said, reasons for why HHHHT was considered least likely fell into two distinct subcategories. The responses of Francine, Gerard, and Hannah exemplify the ratio of heads to tails being used as a clue (e.g., B, it's the only one with one T and four Hs, the rest have 3 Hs and 2 Ts.), which has been seen in prior research (e.g., Kahneman & Tversky, 1974).

Batanero and Serrano's (1999) found "that students had a greater difficulty in recognizing run properties than frequency properties [which] indicates that the similarity between the observed and expected frequencies may be more important than the run lengths in students' deciding whether a sequence is random" (p. 562). As such, a second subcategory of reasons for HHHHT being the least likely sequence to occur used the population ratio, but, arguably, in a more subtle manner because runs involve ratio implicitly.

Sample response justifications for multivalence of the sequence element: Runs.

Dianne: the chances of getting the same one four times is least likely.

Evan: because it is hard to get one consecutive side to be flipped repeatedly.

Uri: B —because it is unlikely that you will flip heads 4 times and then one tails. But you could also say F because anything is possible. But my final answer is B

Dianne mentioned, "the chances of getting the same one four times is least likely." In this instance there is recognition of the ratio of head to tails, but ratio is not, necessarily, at the crux of the explanation. Dianne's comments are related to the length of the run of heads seen in the HHHHT sequence, four heads in a row. While the response is normatively incorrect, it can be inferred from Dianne and Evan's explanations that the interpretation of the sequence element of the task being used to answer the question posed are not the researcher's intended interpretation because the sequence is being interpreted in terms of runs. The interpretation of the sequence via runs and not the ratio of heads to tails demonstrates alternative unintended interpretations, and bolsters Cox and Mouw's (1992) findings that disruption of one aspect of the representativeness heuristic, i.e., the population ratio, did not exclude the other, i.e., the appearance of randomness, being used as a clue and further demonstrates the multivalence of the sequence element.

Sequence element multivalence: Perceived randomness interpretation

The use of length of run based on the ratio of heads to tails found in the responses of Dianne and Evan is connected to the perceived randomness of the sequence. Also, from the responses of Dianne and Evan it is inferred that interpreting the sequence element of the task by the appearance of randomness is also connected to likelihood element. More specifically, sequences that appear more random (i.e., had less of a pattern) are considered to be more likely to occur, and sequences that are considered less random (i.e., were more patterned) were considered less likely to occur, represented in prior research as not being equally representative yet equally likely (e.g., Tversky & Kahneman, 1974). While this connection between the appearance of randomness (via runs) and likelihood is seen in responses for HHHHT as least likely, the connection between randomness and likelihood is also seen in the responses for the sequence HTHTH being least likely to occur, as shown next.

Sample response justifications for HTHTH.

Igor: Usually, when you flip a coin, the answer won't usually be in a pattern. It would most likely be random.

Justine: Because its kind of odd for it to land in a pattern like that. Usually, it's a totally random sequence of heads or tails.

Ken: D because when something is random it doesn't usually go in a pattern.

Research (e.g. Falk, 1981) has shown that randomness was perceived through frequent switches and short runs. It appears from the justifications of Igor, Justine, and Ken, that a perfect alteration of heads and tails—the highest possible number of switches and the smallest possible runs—does not appear random. Further, the sequence HTHTH did not appear random because it is patterned. The appearance of pattern implies a lack of randomness, and the lack of randomness implies that it is less likely to occur. As such, a perfect alteration of heads to tails demotes the likelihood of the pattern causing it to be the least likely to occur.

Summary

In a general sense, the results from Study I indicate the multivalence of the likelihood, experiment, question, and sequence elements of the CLT. The author's hypothesis—that elements of the CLT are multivalent and are interpreted differently than intended—was previously based on speculation. However, through the analysis of results the inferences were bolstered via indications of multivalence. Indicated in Study I was the multivalent nature of the sequence element of the CLT. The response justifications for HHHHT and HTHTH being the least likely sequences to occur demonstrated the appearance of randomness of the sequence element plays a crucial role in determining likelihood. More specifically, the perception of randomness seen in Study I, and shown in prior research, is based upon two particular features: alterations (or switches) between head and tail (or a tail and a head) and the length of run for a particular outcome if, the author contends, all of the sequences contain the same ratio of heads to tails. Further, and as seen in the response of Uri, the notions of switches and runs can be taken into consideration simultaneously. In declaring that "it is unlikely to flip heads 4 times and *then* one tails" the notion of switches and runs are both taken into consideration. Thus, as seen in Study I, the appearance of randomness for the sequence element of the CLT is based for some on switches, runs, and switches and longest run. To explicate the assertion, Study II maintains the ratio of heads to tails. Further, and upon reviewing the results from Study I, the multivalence of the sequence element provided the bulk of the data and will become the focus of investigation for Study II. Thus, and analogously speaking, Study II will be to Study I what the interpretation of the literature was to the literature review.

Discussion

The dominant interpretation of CLT responses found in the psychology literature declares that individuals utilize the major characteristics of (1) the population from which the sequence originated (approximately an even number of heads and tails), and the major characteristics of the process of flipping a fair coin (a resultant random process). As such, the author infers that (and will attempt to determine if) an individual employing an heuristic approach to answering the CLT interprets the multivalent sequence element of the task via: (1) the ratio of heads to tails, or (2) the regularity of the sequence, which is perceived as reflective of a random process by the presence or absence of pattern based on frequent switches and thus short runs (i.e., perceived randomness determinants of short sequences).

The author further infers that from an alternative perspective of the sequence element it is understandable that an individual may be answering a question other than the one they were originally posed and, thus, it becomes understandable that an individual may answer that one of the sequences presented is least likely to occur. Further, and also from this new perspective of the sequence element, it becomes possible for researchers to begin to infer why an individual may answer that one of the sequences presented is least likely to occur.

As seen in the literature, an individual may answer that THTTT is the least likely to occur because of all the sequences presented THTTT's ratio of head to tails is the furthest away from an equal number of heads and tails, which is expected from sample to parent similarity. Interpreting the sequences of outcomes (i.e., sequence element) in terms of the ratio of the number of heads and tails implies a distinguishing of sequences (according to a particular attribute), in this case the ratio of heads to tails. While the ratio of heads to tails interpretation is not the intended interpretation from the individual implementing the task, the act of discerning between sequences is a key element to the task and, furthermore, is implied in the asking of the question. A caveat in this situation is that individuals are right in their attempt to discern between the different sequences, but in the end the 'correct' response requires them to make no discernment at all and indicate that each of the sequences are equally likely to occur.

An individual who considers the major characteristics of the process (and not the population) would, hypothetically, focus their attention on the three sequences that consist of the same ratio of heads to tails (THHTH, HHHTT, HTHTH), and their attention would be focused on the perceived randomness of the sequences (i.e., characteristics of the process) in terms of regularity and the law of small numbers (i.e., short runs and frequent switches). In fact, given that regularity is not reflective of a perceived random process, the sequence of perfect alterations, i.e., HTHTH, may be deemed the least likely to occur, as seen in Study I. Moreover, the sequence HHHTT may also be considered the least likely to occur, as seen in

Study I, because of (1) the apparent regularity, and (2) the lack of frequent switches and thus short runs associated with the law of small numbers. For example, according to a switches interpretation of the sequence element of the task, sequences can be ranked from the lowest number of switches to the highest number of switches (or highest to lowest): HHHTT, THTTT, THHTH, HTHTH. While this particular interpretation does not single out a particular sequence (to be quickly thinned from the herd), the possible "ranking" of sequences may play a role in determining the least likely sequences for different interpretations of the likelihood element.

Taking everything presented in the summary and discussion of Study I into consideration, one goal of Study II is to bolster the author's speculation(s) that respondents are in fact interpreting the sequence element in terms of (1) the number of switches, (2) the length of the longest run, or (3) a combination switches and the longest run. Further, and as mentioned, all main goals of Study II will be focused on the subjective-sample-space.

CHAPTER FOUR

Study II

The bulk of Chapter Four is dedicated to presenting the results from Study II. In fact, the majority of the pages are dedicated to presenting the transcription of 42 out of a possible 239 response justifications. Dedicating the chapter to 42 response justifications has been done deliberately on the part of the author, because, and as presented in the interpretation of the literature in Chapter Two, there is an invariant method witnessed in each iteration of a CLT response interpretation, which always begins with the results. Given the interpretive nature of the research, the author felt it important to have the data to be analysed presented, at some point, in its 'rawest' form. However, before the results are offered, details associated with Study II are presented.

CLT: Ratio control

A variety of issues (e.g., forced response items) have been addressed with subsequent iterations of the CLT. However, one element has remained the same. In each instantiation of previous research, subjects have always been presented sequences with a disparate population ratio in each of their choices. For example, one response option would contain three heads and three tails (e.g., HTHTTH), and the second choice would contain five heads and one tail (e.g., HHHHHT).

While *all* previous research on the CLT (of note, Shaughnessy's (1977) research did contain the same ratio of boys and girls for a most likely, not least likely, version of the task) occurred with sequences with a disparate number of heads and tails, the present study examines responses from respondents when all sequence choices presented contain the same ratio of heads to tails. In doing so, the author, working from the findings of Cox and Mouw (1992)—who found disruption of one aspect of the representativeness heuristic such as the appearance of randomness, did not exclude the population ratio being used as a clue—has deliberately used sequences which all contain the same ratio of heads to tails. Controlling the number of heads and tails in the sequence was done in order to encourage alternative interpretations of the sequences element in terms other than ratio of heads to tails, such as switches and runs.

Based upon the task considerations presented above, students were presented with the following task (which is similar in wording to Konold et al.'s (1993) iteration of the task) shown in Figure 11.

Which of the following sequences is least likely to result from flipping a fair coin five times. Provide reasoning for your response…

(A) H H H T T
(B) H H T T H
(C) T H H H T
(D) T H H T H
(E) H T H T H
(F) All sequences are equally likely to occur

Figure 11. CLT utilised in Study II

Participants

Participants in this study were 239 prospective teachers. More specifically, there were 163 prospective elementary teachers enrolled in a methods for teaching elementary mathematics course, which is a course in the teacher certification program; and 76 prospective secondary teachers enrolled in a methods for teaching secondary mathematics course, also a course in the teacher certification program. The 163 elementary teachers consist of students in five different classes over two different years, taught by two different instructors. The 76 secondary teachers consist of two classes taught by the same instructor in two different years. In all instances, the task was presented prior to the introduction of probability to the course; however, some if not most participants will have encountered probability in their previous schooling.

Results of Study II

Given the task presented to the participants consisted of two major components—answer, and justification—the data from the study is similarly presented.

Answers

While there are six choices to the CLT presented, numerical results, seen in Table 1, can be categorized as falling into one of two categories: normatively correct responses or normatively incorrect responses. Of the 239 participants who took part in the study, approximately 82 percent (197/239) correctly chose that each of the sequences presented were equally likely to occur. For elementary teachers the percentage was approximately 81 percent (132/163), and for secondary teachers the percentage was approximately 86 percent (65/76). However, as presented in interpretation of the literature and demonstrated in Study I, the numbers presented do not necessarily mean all 197 participants correctly arrived at the correct answer. While this discrepancy is worthy of investigation, pursuing such an investigation is not in accordance with the scope of the present research, and will be addressed at the end of the research when a CLT research agenda will be presented by the author.

The 42 participants (roughly eighteen percent) that chose a normatively incorrect answer to the CLT are comprised of 8, 13, and 20 participants who chose HHHTT, THHHT, and HTHTH as least likely, respectively. For elementary teachers 8, 7, and 16 chose HHHTT, THHHT, and HTHTH as least likely, respectively. For secondary teachers 0, 6, and 4 chose HHHTT, THHHT, and HTHTH as least likely, respectively. Also of note, within the five normatively incorrect responses to the task, one participant chose HHTTH least likely and zero participants chose THHTH as least likely.

Classes	Number of participants	(A) HHHTT	(B) HHTTH	(C) THHHT	(D) THHTH	(E) HTHTH	(F) equally likely
Prospective Elementary Teachers (PET)							
Class 1	32	2	0	2	0	1	27
Class 2	28	1	0	2	0	1	24
Class 3	34	0	0	1	0	4	29
Class 4	31	0	0	1	0	4	26
Class 5	38	5	0	1	0	6	26
PET Total	**163**	**8**	**0**	**7**	**0**	**16**	**132**
Prospective Secondary Teachers (PST)							
Class 6	40	0	0	3	0	1	36
Class 7	36	0	1	3	0	3	29
PST Total	**76**	**0**	**1**	**6**	**0**	**4**	**65**
Total	**239**	**8**	**1**	**13**	**0**	**20**	**197**

Table 1. Numerical results of Study II

Justifications

Recognizing the multivalence of CLT responses, as presented in Chapter One, the 41 of the 42 (the HHTTH response has been left out at the author's discretion) response justifications for why HHHTT, THHHT, and HTHTH were chosen least likely are presented below, untreated in any fashion. Response justifications are categorized according to elementary and secondary teacher responses for each of the three sequences. As with Study I, pseudonyms are used for Study II.

Response justifications for HTHTH: Elementary teachers.

Claire: 1st choice: (F) All have the same likelihood of occurring is what I think. It's random. 2nd choice: (E) The chances of a nice tidy pattern like these seems unlikely.

Michael: It's all pretty random but HTHTH seems too perfectly organised.

Sayid: It's hard to find a pattern, so the ones that are the most random are most likely to happen

Sun: I think it's not likely for it to follow a pattern.

Aaron: I think it's [HTHTH] least likely because it has a pattern.

Ana: random flipping does not produce neat patterns like this.

Kate: I believe there is a 50/50 chance that the first flip will be a heads or a tails. Therefore, I believe that E is least likely to occur b/c the odds of flip a coin from heads to tails is fairly slim.

Ben: I think HTHTH is low percent because it appears alternately.

Ethan: In my opinion, 'E' is the least likely occur because it is hard that different sides continually appear.

Penny: it's least likely that every flip will alternate between heads/tails. However, I think all sequences are equally likely to occur.

John: E is least likely to occur because the chances of having the coin land on the opposite side each time to create a pattern of HTHTH are very slim, the longer the pattern the less likely it will be. Also, to get 3 H's in a row [sequence A] is probably next least likely.

Hurley: Although there is a 50% chance of getting a H or a T. It is very unlikely that you can get a sequence of alternating sides randomly. The probability of this sequence happening would be the least likely.

Jack: With E, an alternating sequence could occur but not necessarily in this order, H + T are more likely to occur at a more random interval.

Sawyer: (E) is least likely to occur because with a 50/50 chance it is unlikely that the results will be alternating H/T with each coin flip. It is more likely that the results would be random.

Juliet: I believe that E is the least likely answer because it is too perfect of a pattern. Even though there is a 50/50 chance of the coin coming up heads or tails, it is very unlikely that it would rotate between the two each flip.

Tom: because the others are more random they are more likely, but to alternate 1 and 1 each time no, that's like orchestrated fairness…doesn't happen it has to be guided.

Response justifications for HTHTH: Secondary teachers.

Eko: e is least likely as it is patterned. Patterns are less likely to arise from random events, but a would be the second less likely as there are 3H in a row, but it is not highly unlikely, just more so than b and d

Jin: I would think the odds of getting a perfect HTHTH pattern are slim (at least one letter would be off most of the time)

Rose: If a coin is flipped five times, the chance of it going from head to tails, head to tails…is not likely. Rather, the coin will likely go from tails and head randomly.

Bernard: Although not impossible, I think e (the alternating HTHTH) is least likely to occur b'cuz flipping a coin is a random act and option e is not.

Response Justifications for THHHT: Elementary teachers.

Gary: I think c because flipping a coin is so random that I don't think the probability of having a t at the start and end of your sequence is that likely. But with E, an alternating sequence could occur but not necessarily in this order, H and T are more likely to occur at a more random interval.

Walt: because three of one letter is unlikely.

Alex: I can't remember the last coin I flipped that came out tails first, so I narrowed it down to either A, B, or E being more likely, which left C, and D. In the end, I chose C as least likely to occur because it has 3 heads in a row.

Mikhail: C is least likely to occur, since the probability for H to occur thrice in succession is tougher.

Karl: For time (c) it does not seem probable that you would roll heads three time in a row. Also, for time (a) by the third roll there was a 100% heads rate and that would result from a fair coin flip

Martin: A side of a coin often appears twice. It is not easy to appear the same side three times, but it is more possible that two sides appear, logically.

George: it's hard to have the same symbol occur three times in a row.

Response Justifications for THHHT: Secondary teachers.

Vincent: Although not impossible, I think c is least likely to occur because flipping a coin is a random act, and option c is not.

Edward: Given the probability of 50% chance of either getting heads or tails, getting a particular sequence ex: HHHTT or THHHT is less likely since a more even sequence should occur.

Danny: I believe it is least likely to get H 3 times in a row. Once the first head is flipped there is a less likely chance that another H would follow and an even more likely chance that a third will happen.

Naomi: It's least likely to have 3 heads continually from flipping a fair coin five times.

Frank: It would be very unlikely that you would get heads three times in a row or have it alternate. Although it is possible, it is very unlikely.

Jacob: Any coin toss has a 50% chance to be either heads or tails. Hence, c) is the least likely pattern to occur due to the decreasing chance of getting 3 heads in a row.

Response Justifications for HHHTT: Elementary teachers.

Boone: (A) because getting three in a row of one type is less probable than the other options of alternating or only two in a row.

Shannon: 5 times, and chances are most likely to be 3H 2T or 2H 3Ts. For the sequences they will more likely to be scrambled, because that's fact. I've tried few times, scrambled.

Libby: (A) b/c what are the chances to get three H's in a row, and two T's after that?

Charlie: because it is very unlikely that there will be the 3H in a row and then 2T.

Nikki: I thought that it wasn't likely to be on one side for three flips and then the other for the rest.

Desmond: I found A less likely to occur, because I find 3 consecutive heads is less likely to occur, when compared to HHTTH or THHTH or HTHTH. It also seems unlikely that tails will follow consecutively right after 3 tosses of heads.

Leslie: It's very hard to get 3 heads in a row and then 2 consecutive tails.

Daniel: chances of getting 3 heads then two tails is low

With information garnered from the previous chapters on research involving the CLT—especially the abductive approach used in past and present CLT research model creation—the next chapter explicates the creation, exploration, examination, testing, and critique of hypotheses, generated by the author's inferences made to best explain the data from Study II.

CHAPTER FIVE

Revealing the Subjective-Sample-Space

Study I evidenced that certain individuals were interpreting the sequence element of CLT differently than intended and, consequently, as the author asserted, were also answering an unintended interpretation of the task. In this chapter the author presents a model, framework, and method all used to analyze and explicate interpretations of the sequence element by respondents in Study II. The author—in a contribution to research in mathematics education in general and CLT research in specific—hypothesizes that (1) an individual's *subjective-sample-space* models certain responses of the CLT, (2) the *meta-sample-space* can act as the framework for analysis of individuals' subjective-sample-space, and (3) the method of *event-description-alignment* aligns the model with the framework. Alternatively stated, the meta-sample-space and event-description-alignment are employed to reveal the subjective-sample-space.

Subjective-sample-space

The application of personal theories or informal conceptions is found in many areas of probability education research (e.g., Konold, 1989), including sample space. For example, "to justify the probabilities for the outcomes of dice games, learners construct informal sample spaces" (Speiser & Walter, 2001, p. 61). The inferred structure of personal sample spaces has been used to demonstrate particular anomalies found in probability education research. For example, the research of (Maher, Speiser, Friel, & Konold, 1998; Speiser & Walter, 2001) demonstrated that certain individuals found, for example, the outcome (5,6) for the experiment of rolling two dice to be as likely as the outcome (6,6). Further, researchers extrapolated the lack of discernment between pairs to all outcomes (e.g., (3,4) and (4,3)). Consequently, researchers hypothesized—and then concluded—that the sample space employed by certain individuals answering the question consisted of 21 possible outcomes and not 36 outcomes, because individuals treated the outcome (5,6) and (6,5) as one outcome. For the researchers, responses to the task revealed a certain structure of the personal sample space used when answering the task.

From engaging with the literature above and in previous chapters, the author speculates that certain individuals answer probability tasks (e.g., the CLT) according to a subjective-sample-space, which is partitioned according to their interpretation of the sequence element of the CLT. Further, the author speculates that examination of respondents' (written) verbal description of events, witnessed in their response justifications, provides insight to the structure of an individuals' subjective-sample-space. A purpose of Study II is to test the author's speculations. First, however, the subjective-sample-space will be discussed further.

In general, Study I demonstrated that the elements of the CLT were multivalent. More specifically, certain responses from Study I (e.g., Francine, Gerard, and Hannah) demonstrated that the sequence element of the CLT was interpreted (in an unintended fashion) according to the ratio of heads to tails. For familiarity, the response justifications are once again presented.

Francine: It's most unlikely to have four heads and one tail because there is a 50% chance.

Gerard: It has the "most uneven" amount of heads and tails.

Hannah: B, it's the only one with one T and four Hs, the rest have 3 Hs and 2 Ts.

Hannah's response, for example, demonstrates the interpretation of the sequence element according to the ratio of heads to tails. The sequence HHHHT for Hannah is interpreted as the "one with one T and four Hs." Further, all other sequences presented (HHTTH, THHHT, HTHTH, THHTH) are interpreted with a four to one mapping to "have 3 Hs and 2 Ts." The order-less interpretation of the sequence seen in prior research has also concurrently raised issues of correctness (Abrahamson, 2008). Hannah is incorrect that HHHHT is less likely than any of the other sequences presented. However, if Hannah's interpretation of the sequence element, which dictates her answering of the question is taken into consideration, she is, in fact, correct in declaring that a sequence with one tail and four heads is less likely (half as likely) than a sequence with three heads and two tails, as seen in Table 2; sequences with three heads and two tails are twice as likely to occur than sequences with one tail and four heads.

0H & 5T	1H & 4T	2H & 3T	3H & 2T	4H & 1T	5H & 0T
TTTTT	TTHTT	TTHTH	HHTHT	HTHHH	HHHHH
	TTTTH	TTTHH	HTTHH	HHTHH	
	TTTHT	THTHT	HTHHT	HHHHT	
	HTTTT	HHTTT	**HHTTH**	THHHH	
	THTTT	THTTH	**THHTH**	HHHTH	
		THHTT	**HHHTT**		
		HTTTH	THTHH		
		HTTHT	**THHHT**		
		HTHTT	TTHHH		
		TTHHT	**HTHTH**		
1	5	10	10	5	1

Table 2. Sample space partition: Ratio of heads to tails.

The author infers that Hannah answers the task according to a subjective-sample-space, which, in this instance, is partitioned according to her ratio of heads to tails interpretation of the sequence element of the CLT. Other notions, including pattern and randomness and likelihood, are also at play in Hannah's ratio response to the CLT, and will be examined in detail during the analysis of Study II results in Chapter Six. However, before analysis of results occurs, the author's framework is now explicated.

Meta-sample-space

The author contends that the conventional sample space (i.e., the set of all possible outcomes for an experiment) is a necessary, but insufficient theoretical framework for analysis of responses to the CLT, because often "experimenter and subject will conceptualize different sample spaces or different frames which may provide the impetus for misinterpretation of the data" (Keren, 1984, p. 122). Further, the author contends that the meta-sample-space, defined as the set of all possible sample space partitions (or a sample space of sample spaces) for an experiment, is a more appropriate theoretical framework for analysis of CLT responses.

In order to illustrate and enumerate the aforementioned assertion, consider the experiment where a fair coin is flipped twice. The resulting sample space would be {(H, H), (H, T), (T, H), (T, T)}, which can concurrently be interpreted as a set consisting of four elements. More specifically, the sample space {(H, H), (H, T), (T, H), (T, T)} would correspond to the set, and each of the outcomes (H, H), (H, T), (T, H), and (T, T) correspond to elements of the set. Using the sample space as a set, a Bell number is "the number of ways a set of n elements can be partitioned into nonempty subsets…and is denoted B_n" (Weisstein, 2007). According to the notion of Bell numbers there are fifteen partitions for a set of four elements, which are found in Table 3.

{{(H, H)}, {(H, T)}, {(T, H)}, {(T, T)}}
{{(H, H), (H, T)}, {(T, H), (T, T)}}
{{(H, H), (T, H)}, {(H, T), (T, T)}}
{{(H, T), (T, H)}, {(H, H), (T, T)}}
{{(H, T)}, {(H, H), (T, H), (T, T)}}
{{(T, H)}, {(H, H), (H, T), (T, T)}}
{{(T, T)}, {(H, H), (H, T), (T, H)}}
{{(H, H)}, {(H, T), (T, H), (T, T)}}
{{(H, H)}, {(H, T), (T, H)}, {(T, T)}}
{{(T, H)}, {(H, H), (T, T)}, {(H, T)}}
{{(T, H)}, {(H, H), (H, T)}, {(T, T)}}
{{(H, H)}, {(H, T), (T, T)}, {(T, H)}}
{{(H, T)}, {(H, H), (T, T)}, {(T, H)}}
{{(H, T)}, {(H, H), (T, H)}, {(T, T)}}
{{(H, H), (H, T), (T, H), (T, T)}}

Table 3. Partitions of the sample space for two flips of a coin

Defining the meta-sample-space as the set of all possible sample spaces allows for alternative representations of the set of all possible outcomes. The emboldened set at the top of Table 3 corresponds to the normative representation of the sample space—seen in classical probability and found throughout research in mathematics education—where outcomes are individually partitioned. However, each of the other fourteen partitions presented above also contain all possible outcomes for the experiment. Thus, it can be argued that the other fourteen partitions (unconventionally) represent the sample space, because they also contain all possible outcomes. Consequently, the meta-sample-space for two flips of a coin contains fifteen sample spaces, or sets of all possible outcomes, differently partitioned.

Jones, Langrall, Thornton, & Mogill (1997) state: "research evidence with regard to sample space is conflicting and highlights the need to study...thinking in this construct more comprehensively, and *within a probability context*" (p. 105, this author's italics). Keren (1984) states that, "knowledge of the sample space used by subjects is crucial for understanding their responses" (p. 127). Fortunately, the meta-sample-space can address the raised issues by providing a probabilistic investigative framework for analysing subjective-sample-spaces, and thus be used as an investigative theoretical framework for responses to the CLT. The enumeration of all possible set partitions explicates the multitude of investigative frameworks available to the researcher; however, one caveat for the CLT with five coins flips is that Bell number thirty-two is equal to 128 064 670 049 908 713 818 925 644.

Sample space partitioning itemises possible partitions—and enumerates the meta-sample-space—that could be used to compare the subjective-sample-space provided by respondents. For example, the partition {{(H, H)}, {(H, T), (T, H)}, {(T, T)}} would correspond to a subjective-sample-space organized according to "the number of heads." Awareness of partitions allows for consideration of more than one particular sample space to be applied to the research. Each of the partitions considered may be used on their own, or in unison with each other in order to expose different interpretations. In an experiment involving light, different combinations of light act as a way to investigate colour. A red, blue and green flashlight will provide a number of different perspectives when the lights are singularly, and in combination, shone on an object. In a similar fashion, the meta-sample-space can shed new light on the comparative likelihood task. By shining the normative light (i.e., partition) alone, all that can be determined is whether a subject is correct or incorrect, and there exists little explanatory power. However, combining the lenses of the different partitions of the meta-sample-space for analysis of the comparative likelihood task there may exist a partition according to which the answer can be considered correct. Different combinations of partitions, along with different numbers of partitions, will allow for multiple perspectives and proper investigation to how subjects are answering probability tasks. To illustrate how an individual's subjective-sample-space aligns with a particular partition of the meta-sample-space, the methodology to be used will now be discussed.

Event-description-alignment

While it has been presented that inferences *can* be made about an individual's subjective-sample-space, *how* inferences are to be made about the subjective-sample-space has not been addressed, until now. The author proposes that one possible solution is through the alignment of verbal description of events with the most appropriate set description of events. In other words, to identify the subjective-sample-space within the meta-sample-space through, what the author calls, event-description-alignment.

Consider the following example of the CLT: A fair coin is flipped three times, which of the sequences is least likely to occur (a) *HHH*, (b) *HHT*, or (c) *the sequences are equally likely to occur*? Given the axiomatic approach to probability will act as the foundation for the theoretical framework for analysis of responses to the CLT, some basic elements of probability theory, and associated terminology are reviewed for the answering of this example.

In the example of the CLT presented, the *experiment* (ξ), i.e., the method or procedure to produce the data, is the flipping of a fair coin three times, and *outcomes* correspond to the results of the *experiment*. The *sample space* (Ω), i.e., the set of all *possible outcomes*, of the *experiment* is {{HHH}, {HHT}, {HTH}, {THH}, {TTH}, {THT}, {HTT}, {TTT}}. Moreover, an *event* (*E*) is an individual (e.g., E_1={HTH}), or group (e.g., E_2={HTH, THH, HHT}) of *designated outcomes* of the *experiment*, and the collection of all subsets of the sample space (i.e., events) comprises the event space. Further, the assumption is made that for each event *E* in the sample space (Ω) there exists a value *P(E)*, known as the probability of *E*. Thus, the mathematical object associated to the experiment consists of: the sample space, the event space, and the probability function, which together are known as the probability space. Further, the assumption is made that the probabilities satisfy three particular axioms: $0 \leq P(E) \leq 1$, $P(\Omega)=1$, and the additive property. Assumptions do not stop with axioms. In this particular interpretation of probability theory the assumption is made that each outcome of the sample space is equally likely to occur. This assumption of equally likely outcomes, coupled with axioms presented, further implies assumptions that (1) P({HHH}) = P({HHT}) = … = P({TTT}), and (2) for this particular example $P(\{E_1\})=1/8$ are made. Further, it follows from the third axiom that for an event *E*, the probability of *E* is the proportion of the number of outcomes in the event *E*, *n(E)*, to the number of outcomes in the sample space, $n(\Omega)$. In other words, *P(E)= number of outcomes in E / number of possible outcomes in* Ω*=* $n(E) / n(\Omega)$. For example, the $P(E_1)$ *= one outcome in* E_1 */ eight outcomes in* Ω; thus, the probability of $P(E_1) = 1/8$. Similarly, the $P(E_2)$ *= three outcomes in* E_2 */ eight outcomes in* Ω; thus, the probability of $P(E_2) = 3/8$.

Considering the original question posed in the CLT example, *P(HHH)=1/8*, and *P(HHT)=1/8*; thus, the answer is (c) *the sequences are equally likely to occur* and an incorrect response to the CLT would be to indicate that one of the sequences HHH or HHT is less likely to occur. However, while responses of HHH being less likely to occur are considered in error, this author wishes to assert that the response is not necessarily void of correct probabilistic answering. Furthermore, the argument is put forth that the differences between "correct" and "incorrect" responses to the CLT are nothing more than differences in description of events (i.e., individual or groups of outcomes).

As shown, there exists more than one sample space partition for an experiment. For example, the sample space for three flips of a coin could be Ω={{HHH}, {HHT}, {HTH}, {THH}, {TTH}, {THT}, {HTT}, {TTT}}, or could be Ω_2 ={{HHH}, {{HHT}, {HTH}, {THH}}, {{TTH}, {THT}, {HTT}}, {TTT}}, or could be Ω_3 ={no tails, at least one tail}. In fact, what may appear to be the appropriate sample space for one individual (e.g., a researcher) may not be the most natural sample space for another (e.g., a student). That said, it is important to recognize that certain conditions must be met. More specifically, the listing of outcomes must be exhaustive, and mutually exclusive. While, Ω, Ω_2, and Ω_3 all represent partitions of the sample space, i.e., of the set of all possible outcomes, for three flips of a fair coin, they vary in number of event outcomes: 8, 4, and 2, respectively. This variability in the number of outcomes associated with the sample space has a direct impact on the probabilities calculated.

This ability to list the set of all possible outcomes, but have variability in the number of outcomes in the partitioned sample space is derived from the notion that "many events which are not themselves outcomes or members of a sample space can be described in terms of those outcomes" (Peck, 1970, p. 119). Moreover, these events, or subsets of the sample space, can have both verbal descriptions and they can have set descriptions. For example, a verbal description of "flipping at least two tails" corresponds to the set description of

{{HTT}, {THT}, {TTH}, {TTT}}, and a verbal description of “a run of two” corresponds to the set description of {{HHT}, {THH}, {TTH}, {HTT}}.

Issues associated with disparity between what is described in words with the numerical equivalent in the set notation also arise. For example, the verbal description of “two heads and a tail” corresponds to a set equivalent of {{HHT}, {HTH}, {THH}} containing three elements. However, if an individual were to collapse the verbal description of the set into a set consisting of one element, i.e., “two heads and a tail” corresponding to a set equivalent of {two heads and a tail} containing one element, there would also be a direct impact on the probabilities being calculated.

While the notions presented above can, on their own, have an impact in the calculating of probabilities, combining the notions of (1) describing the sample space with a varying number of outcomes, and (2) confusion between the number of elements associated with verbal and set descriptions only adds further complications to the calculating of probabilities. In order to demonstrate the probabilistic impact of these notions, consider the following example: If a fair coin is flipped three times what is the probability of getting two heads? Two plausible answers to this question—three-eighths and one-quarter—are now discussed.

While, normatively, the correct answer is three-eighths, perhaps it is more natural, in this instance, to answer one-quarter? Let E_7 = {HHT, HTH, THH} be the set description for the verbal description of “two heads” and, as shown, Ω={{HHH}, {HHT}, {HTH}, {THH}, {TTH}, {THT}, {HTT}, {TTT}} is the set description for the sample space. Given that $n(E_7) = 3$, and $n(\Omega) = 8$, it follows that $P(E_7) = 3 / 8$. However, in order to arrive at this particular answer to the question, the verbal description of “two heads” had to be converted to the corresponding set description of {HHT, HTH, THH}.

Conversion from verbal to set description must occur because when an individual, given a verbal description, argues probabilistically within the domain of verbal descriptions their answer is considered incorrect. For example, an individual who is presented with "two heads," i.e., a verbal description of the question presented may have only four outcomes in their sample space corresponding to Ω_4 = {no heads, one head, two heads, three heads}. Moreover, let E_8 = "two heads", or {two heads}. Given that $n(E_8) = 1$, and $n(\Omega_4) = 4$; it follows, that $P(E_8) = 1 / 4$. While this notion of working probabilistically within the domain of verbal description produces, normatively, an incorrect answer, it can also be argued that the question is answered "correctly."

As shown with the answers of one-quarter and three-eighths, "identification of the sample space is extremely important since different sample spaces (of the same problem) may lead to different solutions" (Keren, 1984, p. 122). Nevertheless, in order to lead to the normative solution, individuals, when asked questions with verbal descriptions of events, must convert those verbal descriptions into set descriptions. While the author is aware that p implies q does not necessarily mean that q implies p, the thought of individuals taking set descriptions of events, and converting them to verbal descriptions may be entertained. In other words, and for example, when presented with a sequence in the CLT, and while the representation of that particular sequence closely resembles the set description (e.g., HHH), do participants convert the presented set description into a verbal description of "three heads?" Moreover, could this conversion be related to incorrect responses seen in the CLT?

More specific to the theoretical example, and corresponding to sample space Ω, consider the following events, and their associated probabilities shown in Table 4.

$E_3 = \{HHH\}$	$n(E_3) = 1$	$N(\Omega) = 8$	$P(E_3) = 1 / 8$
$E_4 = \{HHT\}$	$n(E_4) = 1$	$N(\Omega) = 8$	$P(E_4) = 1 / 8$
$E_5 = \{\text{two heads}\}$	$n(E_5) = 3$	$N(\Omega) = 8$	$P(E_5) = 3 / 8$
$E_6 = \{\text{three heads}\}$	$n(E_6) = 1$	$N(\Omega) = 8$	$P(E_6) = 1 / 8$
$E_7 = \{HHT, HTH, THH\}$	$n(E_7) = 3$	$N(\Omega) = 8$	$P(E_7) = 3 / 8$

Table 4. Events and associated probabilities for three flips of a fair coin

The probabilities associated with Table 2 can be used to explain both correct and incorrect responses to the CLT example. For example, $n(E_3)$ equals $n(E_4)$, which is why, normatively, the sequences are equally likely to occur; but, the same argument cannot be said for $n(E_5)$ and $n(E_6)$. In fact, $n(E_5)$ does not equal $n(E_6)$, because while the verbal description (E_6) and the corresponding set description (E_3) have the same number of outcomes (i.e., $n(E_6)$ equals $n(E_3)$), the verbal description (E_5) and the corresponding set description (E_4) do not have the same number of outcomes, i.e., $n(E_5)$ does not equal $n(E_4)$, because the verbal description (E_5) is supposed to correspond to the set description (E_7), i.e., $n(E_5)$ does not equal $n(E_7)$. Thus, verbal and set disparity can be used to show the incorrect response to the CLT that E_3 is less likely than E_4.

As presented, incorrect answers to the CLT can be analyzed by taking into account the perception of different sample spaces. More specifically, the set description, the verbal description, and, moreover, some blending of the two descriptions may be a root cause of different answers for a rather seemingly straightforward probability question, like the CLT. Fortunately, the foundational, axiomatic framework for the analysis for both answers remains invariant. Alternatively stated, despite the fact that there can be more than one perceived sample space for an experiment, the axiomatic approach presented above to look at the responses is unchanged. Thus responses to the CLT can be analyzed axiomatically against a variety of sample spaces instead of just the traditional sample space; and, moreover, the sample space used for analysis can be based upon verbal clues provided by the student.

Theoretical implementation

Revelation of an individual's subjective-sample-space requires connections to occur between the: CLT, respondent, CLT response, event-description-alignment, meta-sample-space, and researcher (who is informed with prior CLT research, literature, and models). What follows is a theoretical account of how the elements, presented above, connect to reveal the subjective sample space. In order to begin, it is assumed that the CLT, CLT response, meta-sample-space, and subjective-sample-space all exist in respective states of superposition.

In a more general sense, the interpretation and answering of the CLT by the respondent was modeled with a subjective-sample-space in a state of superposition; whereas the CLT response was represented in terms of a meta-sample-space, a theoretical framework, also in a state of superposition, used to describe the interpretation of the CLT response by the researcher. The method of event-description-alignment is designed in an attempt to align the subjective-sample-space with the meta-sample-space, that is, the interpretation of the task with the interpretation of the response. Should the two align, there would exist reason to suspect that certain individuals do interpret responses, under conditions of uncertainty, according to a subjective-sample-space.

In a more specific sense: Consider, first, the CLT. Before an individual responds to the CLT, the task is best described as existing in a state of superposition. The answering of the

CLT by the individual is the act, or observation that collapses the state of superposition for the CLT. The goal of this research is, to the best of the author's ability, determine how the task, for certain individuals, was interpreted and answered (i.e., subjective-sample-space), also considered in a state of superposition at the moment. Consider, second, the CLT response given during the respondent's answering of the CLT. Analogous to the subject responding to the CLT and collapsing the task's state of superposition, the author's interpretation of the CLT response collapses the CLT responses' state of superposition. However, and further, considering the verbal description of events through event-description-alignment, and aligning verbal responses with more appropriate set descriptions collapses the state of superposition of the meta-sample-space. Finally, collapsing the meta-sample-space to a particular partition for analysis of responses will, when responses align accordingly, collapse the superposition of the subjective-sample-space revealing which of the many possible subjective-sample-space partitions (if any) were used in the answering of the task. The implementation above has been presented to detail the specifics of a more general goal of the author, which is to interpret the CLT response in order to determine how the subject interpreted the CLT. Given the theoretical implementation has been described, the analysis of results begins.

CHAPTER SIX

Analysis of Study II Results

As evidenced in Study I, certain individuals interpreted particular elements of the CLT differently than intended. Further, and also as evidenced in Study I, certain individuals interpreted the sequence element of the CLT differently than intended. For Study II, 41 of the 239 respondents who chose HTHTH, THHHT, and HHHTT least likely to occur are the focus of the analysis of results. As will be demonstrated, switches, longest run, and switches and longest run attributes influence respondents' unintended interpretations of the sequence element. Analysis of unintended interpretations of the sequence element will utilize the author's meta-sample-space framework and the method of event-description-alignment. In doing so, the author's main hypothesis—certain individuals answer the task according to a subjective-sample-space partitioned according to their interpretation of the sequence element of the CLT—will be tested, which is the purpose of Chapter Six. It will be demonstrated that if the newly hypothesized model were true, the results observed would follow accordingly. As such, there will exist reason to accept the claims of the newly hypothesized model.

For the analysis of results the forty-one justifications for why sequences HTHTH, THHHT, and HHHTT were least likely to occur have been reorganized below. First, responses are categorized to the sequence chosen as least likely. Second, responses are grouped according to three different, yet interrelated, interpretations of responses per least likely sequence. The presence of an *(e)* to the left of an individuals name denotes elementary teachers and, similarly, *(s)* for secondary.

Analysis of HTHTH as least likely

Response justifications for HTHTH

Response justifications (1): pattern v. randomness & likelihood.

(e) Claire: 1st choice: (F) All have the same likelihood of occurring is what I think. It's random. 2nd choice: (E) The chances of a nice tidy pattern like these seems unlikely.

(e) Michael: It's all pretty random but HTHTH seems too perfectly organised.

(e) Sayid: It's hard to find a pattern, so the ones that are the most random are most likely to happen

(e) Sun: I think it's not likely for it to follow a pattern.

(e) Aaron: I think it's [HTHTH] least likely because it has a pattern.

(e) Ana: random flipping does not produce neat patterns like this.

(s) Eko: e is least likely as it is patterned. Patterns are less likely to arise from random events, but a would be the second less likely as there are 3H in a row, but it is not highly unlikely, just more so than a and e

(s) Jin: I would think the odds of getting a perfect HTHTH pattern are slim (at least one letter would be off most of the time)

The underlined portions of the response justifications presented above evidence a connection between pattern versus randomness (read: lack of pattern) and likelihood for all eight individuals. However, individuals' responses, while alluding to randomness, are discussing the appearance or perception of randomness found in the sequences. For example, Aaron's response indicates that HTHTH is least likely because it has a pattern. Further evidenced from the responses: the more patterned (i.e., less random) the sequence the less likely its occurrence, and the less patterned (i.e., more random) the sequence the more likely its occurrence. For example, and according to Sayid, "the ones that are most random [read: least patterned] are most likely to happen." Whereas the other seven responses (e.g., Sun's response: “I think it’s not likely for it to follow a pattern”) demonstrate that a neater or tidier the pattern means the less likely the chances of occurrence. The connection seen between degree of likelihood and appearance of randomness is found in the research literature. For example, Kahneman & Tversky (1972) declared *reflection of randomness* as one of the determinants of representativeness: “The event should also reflect the properties of the uncertain process by which it is generated, that is, it should appear random” (p. 35). In other words, the sequences presented, while being equally likely, were not considered to be equally representative of a random process.

Response justifications (2): ***switches & likelihood****.*

(e) Kate:	I believe there is a 50/50 chance that the first flip will be a heads or a tails. Therefore, I believe that **E is least likely to occur b/c the odds of flip a coin from heads to tails is fairly slim**.
(e) Ben:	I think **HTHTH is low percent because it appears alternately**.
(e) Ethan:	In my opinion, **‘E’ is the least likely occur because it is hard that different sides continually appear**.
(e) Penny:	**it’s least likely that every flip will alternate between heads/tails**. However, I think all sequences are equally likely to occur.

The **emboldened portions** of the response justifications presented above evidence a connection between switches and likelihood. More specifically, each of the four responses evidence that the perfect alteration of heads and tails (e.g., HTHTH), that is the maximum number of switches possible for a sequence, would correspond to the least likely of the sequences to occur. According to Penny, “it’s least likely that every flip will alternate between heads/tails.” The connection between switches and perceived randomness seen in the justifications above aligns with previous research results (e.g., Falk, 1981; Falk & Konold, 1997), which show frequent switches were indicative of the appearance of randomness.

*Response justifications (3): pattern v. randomness & **likelihood** & **switches**.*

(e) John: **E is least likely to occur because the chances of having the coin** land on the opposite side each time to **create a pattern of HTHTH are very slim**, the longer the pattern the less likely it will be. Also, to get 3 H's in a row [sequence A] is probably next least likely.

(e) Hurley: Although there is a 50% chance of getting a H or a T. **It is very unlikely that you can get a sequence of alternating sides randomly**. The probability of this sequence happening would be the least likely.

(e) Jack: With E, an alternating sequence could occur but not necessarily in this order, H + T are more likely to occur at a more random interval.

(e) Sawyer: **(E) is least likely to occur because with a 50/50 chance it is unlikely that the results will be alternating H/T with each coin flip**. It is more likely that the results would be random.

(e) Juliet: **I believe that E is the least likely answer because it is too perfect of a pattern.** Even though there is a 50/50 chance of the coin coming up heads or tails, **it is very unlikely that it would rotate between the two each flip.**

(e) Tom: because the others are more random they are more likely, **but to alternate 1 and 1 each time no, that's like orchestrated fairness…doesn't happen it has to be guided**.

(s) Rose: If a coin is flipped five times, **the chance of it going from head to tails, head to tails…is not likely**. **Rather**, the coin will likely go from tails and head randomly.

(s) Bernard: Although not impossible, **I think e (the alternating HTHTH) is least likely to occur** b'cuz flipping a coin is a random act and option e is not.

Whereas the first set of response justifications evidenced a relationship between the appearance of randomness (determined through pattern or lack thereof) and likelihood, and whereas the second set of response justifications evidenced a relationship between switches and likelihood, the third set of response justifications evidence a relationship between (1) the appearance of randomness (derived from presence or absence of pattern), (2) likelihood, and (3) switches (or alterations). More specifically, the underlined portions above evidence a connection between randomness and likelihood, the **emboldened portions** evidence a connection between switches and likelihood, and the **<u>underlined and emboldened portions</u>** evidence the relationship between appearance of randomness, likelihood, and switches. For example, Hurley declares (in part), "It is very unlikely that you can get a sequence of alternating sides randomly." In Hurley's response, for example, the low likelihood while connected to the perceived absence of randomness due to the pattern is being determined by the alteration of the coin from heads to tails. Further, for all eight responses shown above the likelihood of the sequence is derived from the absence of presence of pattern, also known as perceived randomness. However, the perceived randomness is derived from the alteration or switches of the sequence. As such, and syllogistically, it is contended that the switches attribute of the sequence element is being employed to determine the likelihood of the sequence. Alternatively stated, it is contended that a subjective-sample-space partitioned according to the switches attribute of the sequence, in this instance, is how subjects are interpreting the sequence element of the CLT and subsequently answering the task.

The justifications provided in participants' responses indicate that the subjective-sample-space they are describing corresponds to an entirely different partition than the sample space responses are conventionally and traditionally pitted against. Based upon the verbal descriptions presented, a more appropriate or natural set description, corresponding to the verbal descriptions given, would be to collapse the meta-sample-space (via event-description-alignment) to a sample space partitioned according to switches, as shown in Table 5.

0 Switches	1Switch	2Switches	3Switches	4Switches
HHHHH	**HHHTT**	HHTHH	HHTHT	***HTHTH***
TTTTT	TTTHH	TTHTT	TTHTH	THTHT
	TTHHH	**HHTTH**	HTHHT	
	HHTTT	TTHHT	THTTH	
	HHHHT	HTTHH	**THHTH**	
	TTTTH	THHTT	HTTHT	
	THHHH	HHHTH	THTHH	
	HTTTT	TTTHT	HTHTT	
		HTHHH		
		THTTT		
		THHHT		
		HTTTH		
n(0S) = 2	n(1S) = 8	n(2S) = 12	n(3S) = 8	n(4S) = 2

Table 5. Sample space partition: Switches

After having collapsed the meta-sample-space through event-description-alignment, when the response that HTHTH is least likely is pitted against the switches partition of the sample space, the response is correct in that HTHTH is the least likely sequence to occur, because n(4s)<n(3s)=n(1s)< n(2s). In other words, the event of alternating sides every time does have the least number of outcomes when compared to all of the other sequences and, thus, would be least likely. As such, and through an alternative interpretation, all eight people represented in *Response justifications (3)* can be seen as correctly answering the task. For example, Rose can be interpreted as correct in declaring that "if a coin is flipped five times, **the chance of it going from head to tails, head to tails…is not likely**." While incorrect when pitted against the conventional/normative set description of the sample space, Rose is correct when her response is pitted against the switches partition of the sample space.

Further, the responses from *Response justifications (2)* when pitted against the switches partition of the sample space can also be considered as correct in their answering of the task. Thus, 12 out of 20—the claim for 20 out 20 cannot be made because what the 'pattern' is derived from was not able to be determined in *Response justifications (1)*—response justifications for HTHTH being least likely to occur can be considered to possess correct answering when pitted against the switches partition of the sample space.

While a response is judged as incorrect when the verbal descriptions are compared to the normative set description, the same response is judged as correct and demonstrates normative probabilistic reasoning when the verbal descriptions of the sample space are compared to a more appropriate set description of the verbal description presented. Alternatively stated, if it were the case that individuals are employing a subjective-sample-space partitioned according to the switches attribute of the sequence element, then individuals would think that a perfect alteration for the sequence is least likely to occur. As such, there exists reason to accept the claim that the employment of a subjective-sample-space under conditions of uncertainty, in this instance organized according to switches, is taking place when answering the CLT.

Analysis of THHHT as least likely

Response Justifications for THHHT

Response justifications (1): pattern v. randomness & likelihood.

(s) Vincent: Although not impossible, I think c is least likely to occur because flipping a coin is a random act, and option c is not.

(e) Gary: I think c because flipping a coin is so random that I don't think the probability of having a t at the start and end of your sequence is that likely. But with E, an alternating sequence could occur but not necessarily in this order, H and T are more likely to occur at a more random interval.

(s) Edward: Given the probability of 50% chance of either getting heads or tails, getting a particular sequence ex: HHHTT or THHHT is less likely since a more even sequence should occur.

The <u>underlined portions</u> of the response justifications presented above also evidence a relationship between pattern versus randomness and likelihood. For example, Gary declares, “flipping a coin is so random that I don’t think the probability of having a t at the start and end of your sequence is that likely.” For Gary, having tails bookend the sequence is not perceived as random. As such, and as seen in the literature, the sequence that appears less random for Gary, that is to say more patterned, is considered less likely. Once again, the connection between perceived randomness and likelihood is apparent in the response justifications, where more random (i.e., less patterned) sequences are considered more likely to occur, and less random (i.e., more patterned) sequences are considered less likely to occur. While similar to the findings shown in *Response justifications (1)* for HTHTH least likely to occur, the attributes associated with the appearance of randomness are different for THHHT being considered least likely to occur.

Response justifications (2): ***runs & likelihood.***

(e) Walt: **because three of one letter is unlikely**.

(e) Alex: I can’t remember the last coin I flipped that came out tails first, so I narrowed it down to either A, B, or E being more likely, which left C, and D. In the end, **I chose C as least likely to occur because it has 3 heads in a row**.

(e) Mikhail: **C is least likely to occur, since the probability for H to occur thrice in succession is tougher**.

(e) Karl: **For time (c) it does not seem probable that you would roll heads three time in a row**. Also, for time (a) by the third roll there was a 100% heads rate and that would result from a fair coin flip

(s) Danny: **I believe it is least likely to get H 3 times in a row**. Once the first head is flipped there is a less likely chance that another H would follow and an even more likely chance that a third will happen.

(s) Naomi: **It’s least likely to have 3 heads continually** from flipping a fair coin five times.

(e) Martin: A side of a coin often appears twice. **It is not easy to appear the same side three times**, but it is more possible that two sides appear, logically.

(s) Frank: **It would be very unlikely that you would get heads three times in a row** or have it alternate. Although it is possible, it is very unlikely.

(e) George: **it's hard to have the same symbol occur three times in a row**.

The **emboldened portions** of the nine response justifications presented above evidence a relationship between runs (i.e., same occurrence in a row) and likelihood. Further, there is a relationship between the length of the run and likelihood of occurrence, where longer runs are less likely to occur than shorter runs. For example, as evidenced by all nine responses, but succinctly exemplified by Walt, "three of one letter is unlikely." According to Tversky and Kahneman (1974) "Chance is commonly viewed as a self correcting process in which a deviation in one direction induces a deviation in the opposite direction to restore the equilibrium" (p. 7). From the responses above and based on the length of the sequence presented in the task, the occurrence of two heads in a row appears to induce a wanted deviation in the opposite direction. The presence of three heads in a row impacts, in this instance, certain individuals' perception of randomness.

*Response justification (3): <u>pattern v. randomness</u> & **<u>likelihood</u>** **& runs**.*

(s) Jacob: Any coin toss has a 50% chance to be either heads or tails. Hence, **<u>c) is the least likely pattern to occur due to the decreasing chance of getting 3 heads in a row</u>**.

The **<u>underlined and emboldened</u>** portions of the response above evidence a relationship between (1) pattern versus randomness, (2) likelihood, and (3) runs. Jacob's response shows a connection between pattern and likelihood when declaring, "c) is the least likely pattern to occur," but justifies his assertion by further stating "due to the decreasing chance of getting 3 heads in a row." Thus, Jacob's response shows a connection between perceived randomness and likelihood that is based upon a pattern or lack thereof. However, perceived randomness, in this instance, is derived from the length of run found in the sequence. As such, and syllogistically, it is contended that the longest run attribute of the sequence is being employed to determine the likelihood of the sequence. Alternatively stated, it is contended that a subjective-sample-space based on the longest run attribute of the sequence element, in this instance, is how subjects are interpreting the CLT.

Jacob's justification indicates his subjective-sample-space corresponds to an entirely different partition than the conventional partition. In fact, based upon the verbal justification presented (i.e., through event description alignment), a more appropriate or natural set description, corresponding to the verbal descriptions given, would be to collapse the meta-sample-space according to a longest run partition, as shown in Table 6.

LR5	LR4	LR3	LR2	LR1
HHHHH	HHHHT	**HHHTT**	HHTHT	**HTHTH**
TTTTT	TTTTH	TTTHH	TTHTH	THTHT
	THHHH	TTHHH	HTHHT	
	HTTTT	HHTTT	THTTH	
		HHHTH	**THHTH**	
		TTTHT	HTTHT	
		HTHHH	THTHH	
		THTTT	HTHTT	
		THHHT	HHTHH	
		HTTTH	TTHTT	
			HHTTH	
			TTHHT	
			HTTHH	
			THHTT	
n(LR5) = 2	n(LR4) = 4	N(LR3) = 10	n(LR2) = 14	n(LR1) = 2

Table 6. Sample space partition: Longest run

After having collapsed the meta-sample-space through event-description-alignment, when Jacob's response that THHHT is least likely to occur "due to the decreasing chance of getting 3 heads in a row" is compared to the longest runs partition of the sample space, the response is correct in that as the length of the run increases, the probability of the sequence is less likely (i.e., n(LR5)< n(LR4)< n(LR3)< n(LR2)). More specifically, runs of length three are less likely when compared to runs of length two (i.e., n(LR3)<n(LR2). Further, individuals' responses presented in *Response Justifications (2)* may also be considered correct when pitted against the longest run partition of the sample space. For example, Naomi declares, "It's least likely to have 3 heads continually from flipping a fair coin five times." However, before any further claims are proposed or discussed for the runs partition, the second sequence that contains three heads in a row is examined.

Analysis of HHHTT as least likely: Take one

Response justifications for HHHTT

Response justification (1): pattern v. randomness & likelihood.

> *(e) Shannon*: 5 times, and chances are most likely to be 3H 2T or 2H 3Ts. For the sequences they will more likely to be scrambled, because that's fact. I've tried few times, scrambled.

The underlined portion of the response justification presented above again evidences a connection between pattern versus randomness and likelihood. For example, Shannon considers "scrambled" sequences more likely, which would imply unscrambled (read: ordered) sequences would be considered less likely to occur.

*Response justification (2): **runs & likelihood**.*

(e) Boone: **(A) because getting three in a row of one type is less probable than the other options of alternating or only two in a row**.

The **emboldened portions** of the response justification presented above evidences a connection between runs and likelihood. For Boone, the length of the run is directly related to the likelihood of the sequence. More specifically, for Boone, and as also seen in the response justifications for THHHT, the longer the run the less likely the sequences chances of occurrence.

*Response justifications (3): <u>switches &</u> <u>**likelihood**</u> **& runs**.*

(e) Libby: (A) b/c <u>**what are the chances to get three H's in a row, and two T's after that**</u>?

(e) Charlie: <u>**because it is very unlikely that there will be the 3H in a row and then 2T**</u>.

(e) Nikki: <u>**I thought that it wasn't likely to be on one side for three flips and then the other for the rest**</u>.

(e) Desmond: <u>**I found A less likely to occur, because I find 3 consecutive heads is less likely to occur, when compared to HHTTH or THHTH or HTHTH. It also seems unlikely that tails will follow consecutively right after 3 tosses of heads**</u>.

(e) Leslie: <u>**It's very hard to get 3 heads in a row and then 2 consecutive tails**</u>.

(e) Daniel: <u>**chances of getting 3 heads then two tails is low**</u>

The <u>**underlined and emboldened portions**</u> of the response justifications evidence a relationship between the appearance of randomness, likelihood, and runs. In all six justifications the likelihood of the sequence is being determined by the appearance of randomness, which is, in turn, being determined by the presence or absence of a pattern. However, the perceived randomness, in this instance, is derived from the length of the longest run found in the sequence. The results are consistent with prior research (e.g., Falk, 1981), which declared that the appearance of randomness was derived from short runs (thus frequent switches), which in this instance are of the order two or less. As such, and syllogistically, it is contended that the length of the longest run attribute of the sequence element is being employed to determine the likelihood of the sequence. Alternatively stated, it is contended that a subjective-sample-space partitioned according to the longest run attribute of the sequence, in this instance for THHHT being least likely, is how subjects are interpreting the sequence element of the CLT and subsequently answering the task.

An analysis of the response justifications provided by the subjects who responded HHHTT least likely also indicate that the subjective-sample-space they are describing corresponds to an entirely different sample-space-partition. The verbal descriptions presented imply a more appropriate event-description alignment would be to collapse the meta-sample-space to a length of longest run partition, as shown in Table 7.

LR5	LR4	LR3	LR2	LR1
HHHHH	HHHHT	***HHHTT***	HHTHT	**HTHTH**
TTTTT	TTTTH	TTTHH	TTHTH	THTHT
	THHHH	TTHHH	HTHHT	
	HTTTT	HHTTT	THTTH	
		HHHTH	**THHTH**	
		TTTHT	HTTHT	
		HTHHH	THTHH	
		THTTT	HTHTT	
		THHHT	HHTHH	
		HTTTH	TTHTT	
			HHTTH	
			TTHHT	
			HTTHH	
			THHTT	
n(LR5) = 2	n(LR4) = 4	n(LR3) = 10	n(LR2) = 14	n(LR1) = 2

Table 7. Sample space partition: Longest run

After having collapse the meta-sample-space through event-description-alignment, when the response that HHHTT is least likely, because a run of length three is less likely, is compared to the longest runs partition of the sample space, the response is correct in that as the length of the run increases, the probability of it becomes less likely (i.e., n(LR5)< n(LR4)< n(LR3)< n(LR2). More specific to the CLT presented, runs of length three compared to runs of length two possess less outcomes (i.e., n(LR3)<n(LR2). As such, HHHTT would be considered less likely (but not least n(LR5)=n(LR1)=2) among the sequences presented. Moreover, Boone's response that "three in a row of one type is less probable than the other options of alternating or only two in a row," can be considered correct when compared against the longest runs partition of the sample space. Seen yet again, responses can be judged as correct from this new perspective. Further, Boone's response in *Response justifications (2)* for HHHTT being least likely to occur, when pitted against the longest run sample-space-partition can also be considered correct.

If it were the case that individuals are employing a subjective-sample-space partitioned according to the longest run attribute of the sequence element, then those individuals would in fact answer that a longer run of three is less likely to occur. As such, there exists reason to suspect the employment of a subjective-sample-space under conditions of uncertainty, in this instance organized according to length or run. However, when responses are compared to the longest runs partition, the sequences HHHTT and THHHT should have been considered equally likely.

Analysis of HHHTT as least likely: Take two

"The extent of student's beliefs in the acceptability of multiple, conflicting answers to the same mathematics questions is worthy of further research" (Rubel, 2007, p. 554) Unfortunately, when pitted against the conventional/normative set description of the sample space research, while worthy, is unable to be conducted to the extent required. However, when compared to a set description equivalent, based upon verbal descriptions, responses can be shown to possess very subtle innate probabilistic answering and interpretation, which considers more than one factor at a time. In recognition of the aforementioned assertion, the responses from *Response justification (3)* for HHHTT as least likely to occur are revisited, and differently emboldened.

Response justification (3) revisited: switches & runs & likelihood.

(e) Libby: (A) b/c what are the chances to get **three H's in a row**, and **two T's after that**?

(e) Charlie: because it is very unlikely that there will be the **3H in a row and then 2T**.

(e) Nikki: I thought that it wasn't likely to be on **one side for three flips and then the other for the rest**.

(e) Desmond: I found A less likely to occur, because I find 3 consecutive heads is less likely to occur, when compared to HHTTH or THHTH or HTHTH. **It also seems unlikely that tails will follow consecutively right after 3 tosses of heads**.

(e) Leslie: It's very hard to get **3 heads in a row and then 2 consecutive tails**.

(e) Daniel: chances of getting **3 heads then two tails** is low

The verbal descriptions presented do not appropriately collapse to the normative partition, or to the switches partition, or to the longest runs partition of the meta-sample-space. The verbal descriptions presented appropriately collapse to a much more nuanced partition of the meta-sample-space. In each of the justifications presented, participants allude to switches *and* runs. For example, Leslie declares, "it's very hard to get 3 heads in a row and then 2 consecutive tails." In fact, all six responses refer in some manner to how the runs have switched from one type to the other (e.g., *then , after that, right after*).

As established, all response justifications presented show the likelihood of the sequence is determined by the presence or absence of a pattern, also known as the perceived randomness. However, the perceived randomness, thus far, has been derived from different interpretations of the sequence element of the CLT, including switches and longest runs. According to Konold et al. (1991), "subjects hold multiple frameworks about probability, and subtle differences in situations activate different perspectives [which] can be employed almost simultaneously in the same situation" (p. 360). (For example, in the analyses of THHHT and HHHTT (in take one) as least likely, individuals disregarded the perfect alteration of HTHTH, which, in fact, was least likely, when focusing on the longest runs attribute of the sequences.) The author contends, in take two, that the switches and longest run attribute of the sequence element is being employed to determine the likelihood of the sequence. In other words, it is contended that a subjective-sample-space based on a switches and longest run interpretation of the sequence element is how subjects are answering the CLT. To test the author's assertion of a subjective-sample-space based on switches and longest run, the verbal descriptions presented above are pitted against a collapse of the meta-sample-space (i.e., through event-description-alignment) according to a switches and longest run partition, seen in Table 8.

0 Switch & Longest Run 5	1 Switch & Longest Run 3	1 Switch & Longest Run 4	2 Switch & Longest Run 2	2 Switch & Longest Run 3	3 Switch & Longest Run 2	4 Switch & Longest Run 1
HHHHH	**HHHTT**	HHHHT	HHTHH	HHHTH	HHTHT	**HTHTH**
TTTTT	TTTHH	TTTTH	TTHTT	TTTHT	TTHTH	THTHT
	TTHHH	THHHH	**HHTTH**	HTHHH	HTHHT	
	HHTTT	HTTTT	TTHHT	THTTT	THTTH	
			HTTHH	**THHHT**	**THHTH**	
			THHTT	HTTTH	HTTHT	
					THTHH	
					HTHTT	
n(0S&LR5) = 2	n(1S&LR3) = 4	N(1S&LR4) = 4	n(2S&LR2) = 6	n(2S&LR3) = 6	n(3S&LR2) = 8	n(4S&LR1) = 2

Table 8. Sample space partition: switches and longest run

After having collapsed the meta-sample-space through event-description-alignment, when the response that HHHTT is least likely is compared to the switches and longest runs partition, HHHTT is the least likely sequence (of those sequences containing runs greater than one) to occur. Interestingly, when the sequence HHHTT, which contains a run of three, is compared to the sequence THHHT, which also contains a run of three, are pitted against, through event-description-alignment, the switches and runs partition collapse of the meta-sample-space, subjects can be considered correct in saying that HHHTT is less likely than THHHT. All six responses, which indicate in some fashion that 'the sequence wasn't likely to be on one side for three flips and then the other for the rest,' can be considered correct when compared to the switches and longest runs partition.

Thus, and from an abductive approach, if it were the case that individuals were employing a subjective-sample-space based on the switches and longest run attribute of the sequences element of the CLT, then HHHTT would be considered the least likely to occur. As such, the results have demonstrated that there exists reason to suspect the employment of a subjective-sample-space under conditions of uncertainty, in this instance, organized according to switches and longest run.

There are two main components to the analysis of results seen in Chapter Six. Given the 'inference to best explanation' (i.e., abductive) approach to the research and creation of the subjective-sample-space model, the analysis of results has been tailored to (1) bolster the validity (via results) of the author's inferences concerning unintended interpretations of the sequence element of the CLT, which were previously speculated, and (2) demonstrate via event-description-alignment with the meta-sample-space that subjective-sample-space(s) partitioned according to switches, longest run, and switches and longest run attribute of the sequences element of the CLT 'best' explains the results. Alternatively stated, the results of Chapter Six were presented to *test* the inferences made and *demonstrate* the best explanation (i.e., the subjective-sample-space). For each of the sequences examined, it was the case that if individuals were employing a subjective-sample-space based on the switches, longest run, and switches and longest run attributes of the sequence element of the CLT, then HTHTH, THHHT, and HHHTT (respectively) would be considered the least likely to occur. As such, the analysis of results has demonstrated that there exists reason to suspect the employment of a subjective-sample-space under conditions of uncertainty, in these instances organized according to switches, runs, and switches and longest run: Lending support to the author's main hypothesis, that certain individuals answer the task according to a subjective-sample-space partitioned according to their interpretation of the sequence element of the CLT. Summary and discussion of the results from Study II (and Study I) comprise the final chapter.

CHAPTER SEVEN

Summary and Discussion

Summary

Through a review of the literature it was demonstrated that prior research has shown certain individuals respond to the CLT with (1) normative, (2) heuristic (e.g., Tverksy & Kahneman, 1974), and (3) informal (e.g., Konold, 1989, 1991) models of reasoning under conditions of uncertainty. Through an interpretation of the literature, and then through the application of the CLT Response Interpretation Framework, created by the author, it was demonstrated that prior research on the CLT has focused on the multivalence associated with the CLT responses. Consequently, the author proposed a focal shift in research on the CLT. Instead of investigating the multivalence of CLT responses, the multivalence of CLT was investigated. To aid in the investigation the author created a second theoretical framework, the CLT Task Interpretation Framework, to show that certain individuals interpreted the sequence, likelihood, experiment, and probabilistic perception elements of the CLT in unintended ways. Applying the CLT Task Interpretation Framework to the results of Study I bolstered the author's assertion regarding multivalent elements and laid the groundwork for the sequence element to become the main focus of investigation in Study II.

The author's third framework, the meta-sample-space, was developed and used with a new method, also proposed by the author, event-description-alignment, to demonstrate the nuances of probabilities associated with perceived randomness. Study II, which controlled the ratio of heads to tails for the CLT, showed that certain individuals still respond to the CLT with one sequence considered least likely. As seen in prior research, the perception of randomness played a key role in determining likelihood. More specifically, prior research (e.g., Tverksy & Kahneman, 1974) showed the more patterned a sequence appeared the less random it was considered and thus the less likely it was considered. However, and as demonstrated in the present research, the pattern was based on interpreting the sequence element according to switches, runs, and switches and longest runs; supporting the author's main assertion that certain individuals answer the CLT according to a subjective-sample-space partitioned according to their interpretation of the sequence element of the CLT. Thus, while representativeness determinants (e.g., irregularity, local representativeness) play a role in heuristically determining likelihood via perceived randomness, sequence element attributes (e.g., ratio, switches, runs, switches and runs) play a role in answering the CLT via the employment of a subjective-sample-space.

By recognizing sequence element (attribute) partitions of the sample space, subjects who chose sequence HTHTH, THHHT, and HHHTT are unconventionally, but naturally, interpreting the sequence element according to features in the sequences which are least likely to occur. From the new alternative perspective presented, it can be argued that the students were correctly choosing which of the sequences are least likely to occur. Contextually speaking, the author would be remiss to say that individuals who chose the sequences THHHT, HHHTT and HTHTH as least likely are wrong. Instead, it would be more appropriate to say that the individuals are wrong when comparing their responses to the normative solution. In other words, while the CLT has a caveat that, normatively, each of the sequences is equally likely to occur, an ensuing caveat is that subjects' responses to the CLT are incorrect only when compared to the one particular sample space for which the caveat applies.

Discussion

Certain research found in probability education literature has, at its core, a comparison between observed data and the normatively correct answer. Moreover, there exists a tone of sacrosanctity to the normative solution, despite the existence of inherent issues. For example, “the classical definition is essentially circular, since the idea of ‘equally likely’ is the same as that of ‘with equal probability’ which has not been defined” (Lipschutz & Schiller, 1998, p. 88). Despite these foundational issues, research in probability education continues to embrace the normative solution.

What then are the normatively correct responses to the CLT providing researchers: Are researchers looking for individuals to indicate they recognize that the sequences are equally likely because of tautological argumentation? The answer to the question is no, and is evidenced in responses to the CLT being considered correct by stating *that* the sequences are all equally likely to occur, and are not concerned with the issue as to *why* the sequences are equally likely to occur. The distinction—between that and why—is entirely ignored, because most research in probability education focuses on incorrect responses to the CLT. However, if the focus on research is on the incorrect responses, the framework for the incorrect responses should not be ‘stuck’ on the set representation of the normative sample space. In other words, the framework for analysis of responses must evolve beyond the normative solution.

One approach to this evolution of the framework is to examine alternative sample space partitions of the meta-sample-space, as presented in this study. In order to do so, the sacrosanctity of the normative solution in the CLT must be relaxed. Fortunately, this is possible while still being able to maintain the rigor of a mathematical basis. Instead of hinging research on the antiquated notion of sample space, probability education research can employ the notion of a meta-sample-space.

The field of mathematics has long accepted that the notion of sample space is necessary, but not sufficient, mathematical construct. It is now time for mathematics educators to do the same. As seen in the research presented, recognition of a meta-sample-space unveiled the notion of more than one sample space interpretation existing for an experiment. More importantly, it allowed for more appropriate set descriptions of the verbal descriptions provided by the subjects, and showed that research can, and should, be getting more out of responses to the CLT task; other than the answer is wrong when compared to the normative solution. Unfortunately, "a simple comparison between the normatively correct answer and the observed data has little explanatory power" (Keren, 1984, p.127). But, and perhaps with the new framework presented, the call put forth that "we need to know more about how students <u>do</u> learn to reason probabilistically" (Maher, Speiser, Friel, & Konold, 1998, p.82) can be addressed.

Limitations

The reasoning behind the research presented was abductive in nature. The author, to best explain the evidence if true, hypothesized the subjective-sample-space. To test the hypothesis, the subjective-sample-space was compared to partition collapses of the meta-sample-space through the act of event-description-alignment. The analysis of results demonstrated that if students answered according to a subjective-sample-space the results would follow, which they did. Therefore there exists reason to suspect individuals answer the task according to a subjective-sample-space. However, when abduction is used as a mode of reasoning, minor premises, such as the subjective-sample-space, cannot be declared with certainty. As such, despite conclusions presented, none of the assertions made in this research can be declared with certainty.

For the author all probability is conditional, and having all probability be conditional means that the likelihood of the subjective-sample-space is entirely dependent on further research. As such, having engaged with the CLT literature for an extended period of time, and as is customary, the author wishes to present a CLT 'research agenda', which may directly impact the likelihood of the subjective-sample-space.

Research agenda

The author, taking liberties with his involvement with the CLT literature during the research presented, has compiled a list of potential areas of investigation for the CLT over the next number of years.

Catching up with the times

First, CLT research in mathematics education must further mine the seminal works of psychologists Amos Tversky and Daniel Kahneman. However, researchers must do so without being ‘stuck’ in the 1970’s. While research in mathematics education continues to make use of the representativeness heuristic, psychology research has evolved from heuristics and biases to *attribute substitution* and *System I and System II* reasoning (Kahneman & Frederick, 2002). As Shaughnessy did in the late 1970’s, researchers in mathematics education need to bridge some of the psychological research that has occurred over the past 30 years.

Equiprobability: Primary or secondary intuition

Second, when an individual incorrectly arrives at a correct answer to the CLT, are they employing a primary or secondary intuition (Fischbein, 1987)? Coupling the CLT with other probabilistic tasks (e.g., in a family of three children what are the chances of having two girls?) may help in determining if equiprobability is predominantly a primary or secondary intuition; and, concurrently, examine previous questions of validity (Hirsch & O’Donnell, 2001) for the task.

The frequentist and subjective CLT

Third, the author recommends, given the burgeoning role of alternative interpretations of probability (Jones, Langrall, & Mooney, 2007), that versions of the CLT which account for non-classical interpretations of probability are implemented. For example, respondents are first asked to flip a coin a number of times and then determine which sequence is least likely (i.e., frequentist CLT) or asked to 'make up' a number of different sequences and then determine which is least likely (i.e., subjective CLT).

Correlating with the probability of alteration

Fourth, Falk and Konold (1991) demonstrated that individuals interpreted the order of sequences according to a probability of alteration. The author recommends that the probability of alteration is contrasted with the likelihood of certain sequences to see if a direct correlation between likelihood and probability of alteration exists.

Changing the experiment

Fifth, up until now research has involved either flips of a fair coin or the birth of boys and girls. The author recommends different experiments of the CLT be investigated. For example, experiments where the probability of success is not equal to the probability of failure (i.e., non-binomial experiments with dice). In a general sense, the author calls for a 'dePlatonification' of the CLT (e.g., an answer key CLT).

Research contributions

The research presented has contributed to mathematics education in a variety of ways. Beyond contributions detailed in the summary and discussions, particular contributions and brief descriptions are presented in Table 9.

Domain of contribution	Name	Description
Research method	*Multivalentology*	Focused on part the abductive process which involves the exploration, examination, critique, creation, and testing of (i.e., the study of) hypotheses generated via inference when individuals engage with multivalence. The process involves: statement or task, multivalence, superposition, observation, collapse, comment, inferences, and intended interpretation.
Framework	*Response Interpretation Framework (RIF)*	Framework created and employed by the author as a tool to interpret the multivalence elements of CLT response interpretations found in the review of the literature.
Framework	*Task Interpretation Framework (TIF)*	Framework created and employed by the author as a tool to interpret CLT responses from the perspective of the multivalent elements of the CLT.
Methodological	*Ratio Controlled CLT*	The least likely version of the CLT had, for the first time, the same ratio of heads to tails in all of the sequences presented (3:2).
Framework	*Meta-Sample-Space*	Framework created by the author and defined as the set of all possible sample spaces partitions (or a sample space of sample spaces). Employed to align verbal descriptions of events more appropriately with corresponding set descriptions.
Method	*Event-Description-Alignment*	The process of aligning verbal description of events with appropriate set description of events. In other words, aligning the subjective-sample-space with the meta-sample-space.
Theoretical & Cognitive Model	*Subjective-sample-space*	Model of how the sample space is organized or partitioned for certain individuals in their answering of the CLT.

Table 9. Contributions.

Despite the contributions presented, the research agenda, also presented, may eventually impact the likelihood of the subjective-sample-space. In fact, it is quite possible that other research will, at some point, have a significant impact on the likelihood of the subjective-sample-space. The author admits it is contextually fitting to have the conclusions of his research be discussed in terms of likelihood, not with certainty; and, further, looks forward to future investigations of probabilistic relativism where multiple viewpoints represent the subject in a greater context.

References

Abrahamson, D. (2008). Bridging theory: Activities designed to support the grounding of outcome-based combinatorial analysis in event-based intuitive judgment-A case study. In M. Borovcnik & D. Pratt (Eds. of Topic Study Group 13), *Proceedings of the 11th International Conference on Mathematics Education (ICME)*. Monterrey: Mexico.

Ayton, P., Hunt, A. J., & Wright, G. (1991). Randomness and reality. *Journal of Behavioral Decision Making, 4*, 222-226.

Batanero, C., Green, D. R., & Serrano, L. R. (1998). Randomness, its meaning and educational implications. *International Journal of Mathematical Education in Science and Technology, 29*(1), 113-123.

Batanero, C., & Serrano, L. (1999). The meaning of randomness for secondary school students. *Journal for Research in Mathematics Education, 30*(5), 558-567.

Borovcnik, M., & Bentz, H. (1991). Empirical research in understanding probability. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: Probability in education* (pp. 73-106). Dorecht, The Netherlands: Kluwer.

Bennett, D. J. (1998). *Randomness.* Cambridge, MA: Harvard University Press.

Chernoff, E. J. (2008). The state of probability measurement in mathematics education: A first approximation. *Philosophy of Mathematics Education Journal, 23.*

Cox, C., & Mouw, J. T. (1992). Disruption of the representativeness heuristic: Can we be perturbed into using correct probabilistic reasoning? *Educational Studies in Mathematics, 23*(2)*,* 163-178.

Falk, R. (1981). The perception of randomness. In *Proceedings of the fifth conference of the International Group for the Psychology of Mathematics Education* (pp. 222-229). Grenoble, France: University of Grenoble.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgement. *Psychological Review, 104*(2), 310-318.

Fischbein, E. (1987). *Intuition in science and mathematics*. Dordecht, The Netherlands: Reidel.

Garfield, J. B., & Ahlgren, A. (1988) Difficulties in learning basic concepts in probability and statistics: Implication for research. *Journal for Research in Mathematics Education, 19*, 44-63.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, U.K.: Cambridge University Press.

Green, D. R. (1983). A survey of probability concepts in 3000 pupils aged 11-16 years. In D. R. Grey, P. Holmes, V. Barnett, & G. M. Constable (Eds.), *Proceedings of the first international conference on teaching statistics* (pp. 766-783). Sheffield, UK: Teaching Statistics Trust.

Green, D. R. (1988). Children's understanding of randomness: Report of a survey of 1600 children aged 7-11 years. In R. Davidson & J. Swift (Eds.), *The Proceedings of* *the Second International Conference on Teaching Statistics*. Victoria, BC: University of Victoria.

Hirsch, L. S., & O'Donnell, A. M. (2001). Representativeness in statistical reasoning: Identifying and assessing misconceptions. *Journal of Statistics Education*, *9*(2). [Online: http://www.amstat.org/publications/jse/v9n2/hirsch.html]

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: Responding to classroom realties. In F. K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning,* (pp. 909-955). New York: Macmillan.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in instruction. *Educational Studies in Mathematics, 32*, 101-125.

Jones, G. A., & Thornton, C. A. (2005). An overview of research into the learning and teaching of probability. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 65-92). New York: Springer.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*, (pp. 49-81). Cambridge: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and biases.* Cambridge, U.K.: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430-454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237-251.

Keren, G. (1984). On the importance of identifying the correct 'problem space'. C*ognition, 16*, 121-128.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction, 6*(1)*,* 59-98.

Konold, C. (1991). Understanding students' beliefs about probability. In E. Von Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 139-156), Dordrecht, The Netherlands: Kluwer.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, 3*(1)*. [Online:http://www.amstat.org/publications/jse/v3n1/konold.html]

Konold, C., Pollatsek, A., Well, A., Hendrickson, J., & Lipson, A. (1991). The origin of inconsistencies in probabilistic reasoning of novices. In David Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (Vol. 1, pp. 357-362). Voorburg, The Netherlands: International Statistical Institute.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education, 24*(5), 392-414.

Lecoutre, M-P. (1992). Cognitive models and problem spaces in “purely random” situations. *Educational Studies in Mathematics, 23*(6), 557-569.

Lecoutre, M. P., Rovira, K., Lecoutre, B., & Poitevineau, J. (2006). People’s intuitions about randomness and probability: An empirical study. *Statistics Education Research Journal, 5*(1), 20-35. [Online: http://www.stat.auckland.ac.nz/serj]

Lester, F. K. (2007). *Second Handbook of Research on Mathematics Teaching and Learning*. New York: Macmillan.

Lipschutz, S., & Schiller, J. (1998). *Introduction to probability and statistics*. New York: McGraw-Hill.

Lipton, P. (1991). *Inference to best explanation*. New York: Routledge.

Liu, Y. & Thompson, P. W. (2007). Teachers’ understandings of probability. *Cognition and Instruction, 25*(2), 113-160.

Maher, C. A., Speiser, R., Friel, S., & Konold, C. (1998). Learning to reason probabilistically. In S. B. Berenson, K. R. Dawkins, M. Blanton, W. N. Coulombe, J. Kolb, K. Norwood, L. Stiff (Eds.), *Proceedings of the 20th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.* Vol. X. (pp. 82-87). Raleigh: NC.

Paulos, J. A. (1980). *Mathematics and Humor*. Chicago, University of Chicago Press.

Peck, L. C. (1970). *Basic Mathematics for Management and Economics.* Glenview, Illinois: Scott, Foresman and Company.

Peirce, C. S. (1931). Principles of philosophy. In C. Hartshorne, & P. Weiss (Eds.), Collected Papaers of Charles Sanders Peirce (v. 1). Cambridge, MA: Harvard University Press.

Rubel, L. H. (2006). Students' probabilistic thinking revealed: The case of coin tosses. In G. F. Burrill & P. C. Elliott (Eds.), *Thinking and Reasoning with Data and Chance: Sixty-eighth yearbook* (pp. 49-60). Reston, VA: National Council of Teachers of Mathematics.

Rubel, L. H. (2007). Middle school and high school students' probabilistic reasoning on coin tasks. *Journal for Research in Mathematics Education, 38*(5), 531-556.

Schilling, M. F. (1990). The longest run of heads. *College Mathematics Journal, 21*(3), 196-207

Shaughnessy, J. M. (1977). Misconceptions of probability: An experiment with a small-group, activity-based, model building approach to introductory probability at the college level. *Educational Studies in Mathematics, 8,* 285-316.

Shaughnessy, J. M. (1981). Misconceptions of probability: From systematic errors to systematic experiments and decisions. In A. Schulte (Ed.), *Teaching Statistics and Probability: Yearbook of the National Council of Teachers of Mathematics* (pp. 90-100). Reston, VA: NCTM.

Shaughnessy, J. M. (1992). Research in probability and statistics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: Macmillan.

Shaughnessy, J. M. (2003). Research on students' understanding of probability. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 216-226). Reston, VA: National Council of Teachers of Mathematics.

Tarr, J. E., & Lanin, J. K. (2005). How can teachers build notions of conditional probability and independence? In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 215-238). New York: Springer.

Speiser, B., & Walter, C. (2001). Two dice, two sample spaces. In L. Pereira-Mendoza (Ed.), *Proceedings of the Fifth International Conference on Teaching Statistics,* (pp. 61-66). Singapore: International Statistical Institute.

Toohey, P. G. (1995). Adolescent perceptions of the concept of randomness. *Unpublished master's thesis*, The University of Waikato, New Zealand.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small number. *Psychological Bulletin, 76,* 105-110.

Tversky, A., & Kahneman, D. (1973a). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207-232.

Tversky, A., & Kahneman, D. (1973b). Response-induced reversals of preference in gambling: An extended replication in Las Vegas. *Journal of Experimental Psychology, 101,* 16-20.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Tversky, A., & Kahneman, D. (1980). Casual schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology*. Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1982a). Casual schemas in judgments under uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 117-128). New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1982b). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153-160). New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1982c). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293-315.

Weisstein, E. W. (2007) "Bell Number." From MathWorld--A Wolfram Web Resource. [Online: http://mathworld.wolfram.com/BellNumber.html]