2020

Random Sequence Perception Amongst Finance and Accounting Personnel: Can We Measure Illusion Of Control, A Type I Error, or Illusion Of Chaos, A Type II Error?

Keith I. Taylor
Rollins College, kitaylor@rollins.edu

RANDOM SEQUENCE PERCEPTION AMONGST FINANCE AND ACCOUNTING PERSONNEL: CAN WE MEASURE ILLUSION OF CONTROL, A TYPE 1 ERROR, OR ILLUSION OF CHAOS, A TYPE II ERROR?

By

Keith I. Taylor

A Dissertation

Presented in Partial Fulfillment of Requirements for the

Degree of

Executive Doctor of Business Administration

in the

Crummer Graduate School of Business, Rollins College

Winter Park, Florida

2020

Signature Page

Dissertation Defense: 31 August 2020

DBA Candidate: Keith Taylor

The content and format of the dissertation are appropriate and acceptable for the awarding of the degree of Doctor of Business Administration

Committee Chair

Halil Kiymaz, Ph.D.: ____________________________

Committee Second

Nana Amoah, Ph.D.: ____________________________

Acknowledgments

To my living family and my deceased ancestors through whom I received the wandering spirit that made this effort possible.

To my faculty chair, Dr. Halil Kiymaz; second, Dr. Nana Amoah; reader, Dr. Koray Simsek; technical advisor, Dr. Shalini Gopalkrishnan; writing coach, Christopher Posey;

and the 2017-2020 Crummer Rollins faculty.

All pilot respondents; Qualtrics personnel, panels, and respondents.

Lisa Czapka, Duncan McCulloch,

Teresa Lovett,

and to those whose inspirational ideas facilitated this work.

"To understand the misperception of randomness is to gain power, not over fate, but over ourselves." Poundstone (2014)

"… living with delusion that we don't even realize." Silver (2012)

"'Is chance fair?' was a children's question." Bennett (1998)

"Perceiving events as random or non-random has significance for the conduct of human affairs, since matters of consequence may depend on it." Bar-Hillel and Wagenaar (1991)

"Our life is a series of purposeful random events." Beno Karacsony

Abstract

The purpose of this dissertation was to determine if finance and accounting personnel could distinguish between random and non-random time-series strings and to determine what types of errors they would make. These individuals averaging 13 years of experience were unable to distinguish non-random patterns from random strings in an assessment composed of statistical process control (SPC) charts. Respondents scored no better than guessing which was also assessed with a series of true-false questions. Neither over-alternation (oscillation) nor under-alternation (trend) strategies were able to predict type I or type II error rates, i.e. illusion of control or illusion of chaos. Latent class analysis methods within partial least squares structural equation modeling (PLS-SEM) were successful in uncovering segments or groups of respondents with large explained variance and significant path models. Relationships between desirability of control, personal fear of invalidity, and error rates were more varied than expected. Yet, some segments tended to illusion of control while others to illusion of chaos. Similar effects were also observed when substituting a true-false guessing assessment for the SPC assessment with some loss of explained variance and weaker path coefficients. Respondents also provided their perceptions and thoughts of randomness for both SPC and true-false assessments.

Keywords: random sequence perception, time series, statistical process control charts, randomness, patterns, guessing strategies, behavioral finance, Nelson’s rules, finance, accounting, quality improvement process, partial least squares structural equation modeling, desirability of control, personal fear of invalidity, illusion of control, illusion of chaos.

Table of Contents

List of Tables

List of Figures

Figure | Page

CHAPTER 1 – INTRODUCTION

The study of randomness is not a recent undertaking. The ideas of determinism and free will have been chronicled throughout human history. Randomizers from antiquity have been replaced by random number generators of today. The various studies of games of chance have led to modern probability theory. Science has also explored tacit human expertise and the folly of our faulty heuristics and biases. Randomness has yielded to the self-similar fractals of Mandelbrot (1977, 2004), chaos theory of Lorenz (1963), or even the adaption of Bose-Einstein dynamics to cinema outcomes by De Vany (2004). Similarly, randomness and the error types, generated from human interpretation, cross many diverse domains: clinical trials and diagnostics in medicine; law enforcement and judiciary processes in the legal domain; climate change analytics; artificial intelligence such as IBM's Big Blue; or the CEO effect, to name a few.

A recent google scholar search of "heuristics and biases" returned over 172,000 results, excluding citations and patents, while "behavioral economics" returned almost 2.6M instances using the same criteria. Likewise, the 2019 results of the former were 7,200 and of the latter were 27,200, which indicates a continued and robust interest in many scientific fields. Costa, de Melo Carvalho, and de Melo Moreira (2019) have documented this exponential growth in behavioral economics/finance/accounting research between 1967 and 2016, which included 2,019 articles for behavioral economics, 565 for behavioral finance and 69 for behavioral accounting. The authors concluded that this field of study was important, and in a graph, they showed the exponential growth of published articles over almost 50 years.

In this literature search, Ackert and Deaves (2011) were the only authors to put forth this effort in a book format to summarize and synthesize the vast number of studies in various domains of behavioral economics and finance. Similarly, it was discovered that within the psychological research studies and experiments, many of the subject pools were composed of university-level students or, in minor cases, investors at large. Experiments or surveys, whose participants worked in the finance and accounting industry, were rare. Examples of this research are Biggs and Wild (1985), who studied bias in accountants' perception of unaudited results, and Fenton-O'Creevy, Nicholson, Soane, and Willman (2003) who examined traders' behavior, illusion of control, and performance. Additional accounting examples include Uecker and Kinney (1977) and Johnson (1994).

Additionally, within this literature, two studies offered specific challenges for randomness and error study. First, Nickerson (2002) challenged future researchers to allow individuals to show their knowledge or intuitions about randomness directly. Moreover, Fenton-O'Creevy et al. (2003) again requested a better method to identify and manage the illusion of control among stock market traders generally. They suggested that an assessment technique was needed to better understand candidates who might be subject to this behavior.

Furthermore, the literature examining type II error rates is sparse. Two examples included Harris and Osman (2012), who reviewed decisions from a Bayesian perspective and discussed the illusion of control, a type I error, and the illusion of chaos, a type II error. Also, Tellis (2013) analyzed various firm innovation decisions from both error perspectives.

This study proposes a remedy for these sparse findings and researcher requests through the following contributions. First, statistical process control (SPC) charts are, by design, separating random and non-random sequences. So, a diagnostic test has been constructed to assess knowledge and error rates among a time-series or strings of events based on a scenario of stock returns as a response to Fenton-O'Creevy et al. (2003), who showed that illusion of control was a maladaptive behavior. Second, this instrument equally assesses type I and type II error rates, addressing the large research imbalance. Third, to evaluate respondent knowledge and strategies, open-ended questions allowed individuals to detail their thought processes in response to the request by Nickerson (2002), who proposed that subjects should not only be evaluated by psychological experiments but also have the opportunity to explain their conceptions of randomness. Fourth, to address the shortage of participants within the accounting and finance field, industry professionals were selected. Fifth, structural equation modeling (SEM) was used to link this assessment tool to two previously validated psychological scales and other behavioral characteristics. It is believed, neither SPC charts have been used in random perception assessments nor has SEM modeling been utilized in behavioral finance studies. Sixth, given that SPC charts follow a set of non-random rules that have been determined probabilistically, this insight facilitates a transition from a heuristic and biases framework towards a naturalistic decision-making process through the application of those rules. And lastly, knowing these rules for non-randomness also applies to many time-series events, and therefore, contributes to an increasingly universal viewpoint even though these tools have only been used extensively in manufacturing and to a lesser extent in a hospital or other service settings.

This dissertation follows a thematic organization where studies are grouped into topics for explanation and discussion, which is similar to Bar-Hillel and Wagenaar (1991) and Ackert and Deaves (2011). The literature review in chapter 2 begins with a history of randomness and its use in antiquity. In an attempt to understand his/her environment, humanity has attempted to identify repetitions. The desire to develop this control has a rich history that has touched on both the spiritual and the scientific. As randomness can be described as a series of events without patterns, theories of human pattern recognition are briefly discussed. How our senses perceive and interpret patterns will affect classifications, memory, and future recall of those same patterns. Repeated errors based on those classifications can cause repeated failures. Yet, there are conditions that also favor the abilities of humans to use tacit knowledge in pattern recognition for positive outcomes, i.e., recognition primed decision making. In this case, research suggests that conditions exist where this type of knowledge can be successfully applied. Subsequently, both random sequence perception and cognition of randomness are considered as constructs and are differentiated. Much of this research, however, describes the absence of a proper determination of randomness, which leads to human error. The most common error and the one most universally studied is the type I error, a false positive, or seeing a pattern where none exists. Research in human psychology has classified many constructs related to heuristics and biases. The common constructs such as the hot hand fallacy, gamblers' fallacy, and local representativeness are well known common knowledge. Less known errors such as the idea of innate random generators, studies of randomness, memory or learning, and real-life experiences are outlined in detail as well the more insidious behaviors of the illusion of control or illusory perceptions. While psychology and behavioral economics have, for decades, explored the vast majority of these phenomena, their application and knowledge outside these areas are limited. Furthermore, while some scientific studies may diminish in relevance with time, even early behavioral studies remain applicable today.

Conversely, individuals may not observe an actual phenomenon, which creates an opportunity for a false negative or a type II error. Few studies have examined this classification citing specifically a type II error. Some similar discussions can be found in strategic management literature which has documented decreasing firm longevity over the last 50 years, and which has studied the oldest firms, the Henokiens (for longevity see Mauboussin, Callahan, & Majd [2017] or Khan, Ghayas, & Kashif [2019]; for the oldest firms see Funabashi [2009] or Bennedsen & Henry [2016]).

The literature review continues with one section, which focuses explicitly on behavioral finance and accounting. These selected studies examined financial scenarios showing heuristics and biases by both investors and industry personnel that may lead to suboptimal and maladaptive decisions. Often, because finance and accounting are quantitatively oriented, the business assumptions associated with its reliability are taken for granted. Research suggests caution. Afterward, an auditing subsection reviews type I and type II errors and their avoidance within the corporate compliance domain. Interestingly, the most extensive study of both type I and type II errors is found here. Yet, even with extensive knowledge of their impact, susceptibility to those same errors may remain problematic.

The ensuing section expands the discussion of randomness and errors into other domains to obtain a broader vision of its application and potential interactions. For example, after doing some risk consulting for a large casino in Las Vegas, Taleb (2007) observed that three possible catastrophic events had nearly bankrupted the company, and those events were not even considered in the firm's risk model. Furthermore, seeing these same errors in other fields of study accentuates the realization of the limits of human cognition based on senses alone. And if doubt about the importance of the study of randomness and error types still exists, the costs and consequences of the failure to make correct assessments are enumerated. In business, opportunities for great wealth and success are juxtaposed to high human costs and failures.

To conclude Chapter 2, the literature related to all the operational variables incorporated into this study is examined and contended. Measuring randomness with statistical process control charts and the use of Nelson's rules to separate random variation from special cause variation are explored in-depth as well as the weaknesses of the instrument. This universal tool, not readily used outside of quality improvement, provides an opportunity for assessment of both processes and individual perceptions of those processes. Two well-known psychological scales and several inherent human characteristics have been identified that may separate behavior tendencies into type I and type II error propensities. These are explored as fitness for use. For example, conventional wisdom might dictate that payoff curves indicating gains with small losses, first to market, or "look what I did" action strategies are optimal courses of action. Or does research suggest that they may be type I errors? And finally, a method of creating and assessing individual groups is considered and claimed as a method to avoid committing the same type I error that this study is attempting to examine. Cohen's Kappa, as a measure of interrater reliability, incorporates this idea of random chance pairings within its calculation.

In Chapter 3, the methodology is investigated and explained. First, the structural equation model for this endeavor is described as are the construct relationships and proposed hypotheses. Next, the research method is clarified, including its suitability for use in this study. A discussion ensues concerning the sampling method of the selected population along with particular screening measures. Then, data collection and sample size are stipulated. The measures for this study are later designated and classified as either endogenous, exogenous, grouping, demographic, or qualitative constructs. Each measure elicitation includes its scales, sample questions, reliability, and convergent and discriminant validities if applicable. Finally, the analysis plan reviews all steps for the quantitative evaluation of the SEM model, paths, and measures, and the broad phases for the qualitative portion of the study are delineated to evaluate two open-ended questions.

In Chapter 4, the data and analysis are presented. Several pilot studies are described where the entire survey instrument was tested, as were specific methods to the approach of the qualitative questions. The final survey is discussed and includes suggestions from the Qualtrics personnel who assisted in the data collection. Once the survey was active, the data from Qualtrics was collected, reviewed, and approved, but not without some delays. The qualitative portion of the study is discussed in detail before any quantitative structural equation modeling was performed. Next, calculating the grouping variable, Cohen's Kappa, was proven to be problematic, so an alternative is presented. With all the data separated into tranches, the measurement model is explained and is followed with the analysis of the structural model using the partial least squares method. One by one, each hypothesis is shown to be supported or not with the hypothesized path coefficients or by analyses of differences. Procedures for unobserved heterogeneity are also explained, and these techniques uncovered segments whose individuals might have tendencies to the illusion of control or illusion of chaos. Because of certain results to be developed later, the true-false assessment was also used as an endogenous variable. Identically, this model is also tested for unobserved heterogeneity. The chapter ends with a discussion of the findings.

In Chapter 5, several summary conclusions from the research results are outlined along with several ideas for future projects, and as with all research projects, limitations, and improvements to this study are also suggested.

CHAPTER 2 – LITERATURE REVIEW

Before embarking on the history of randomness and pattern identification, an important observation is offered. Zbilut (2004) noted that science has traditionally attempted to separate order from randomness, which was considered a "nuisance." He commented that perhaps a better perspective would be to consider randomness as "an active process which informs order and vice versa" (p. 6). This idea of patterned-randomness or random-patterns is profound. For example, among animals, the zebra and giraffe exhibit similar designs between animals but are also a unique design specific to each animal. Likewise, in humans, fingerprints and DNA exhibit similar uniqueness as a product of informed order through a randomized process. The best of human nature comes from the recognition of this uniqueness of patterned-randomness, while the worst of humanity comes from a superficial recognition of patterns and stereotypes.

Chapter 2 begins with a brief history of randomness, pattern recognition theories, and decision making. Next, randomness and error types principally from the literature of psychology follows. Several articles concerning behavioral finance and accounting are presented. After these, randomness and error types from a wide variety of domains are offered enhancing the impact of the consequences and costs of these errors. A summary follows a detailed discussion of all operational terms.

History

Bennett (1998) recounted a lengthy history of the development of randomizers and divination in antiquity, describing early games of chance, and outlining the advancement of probabilistic thinking. The author noted that instruments of randomness had fulfilled many early functions in strategic choice, division of property, assigning responsibility, or the settling of grievances. The earliest dice games date to 2750 B.C.E in Mesopotamia; another early example was found in Egypt in 1320 B.C.E. Early Greek and Roman gambling games were played with astragali or actual animal bones. Similarly, in Asia, one of the earliest written accounts of the use of chance devices occurred in a Vedic poem written in Sanskrit around 1000 B.C.E. These early devices and their outcomes were often attributed to gods (see also Yronwode [2012] for a comprehensive treatise on shamanic traditions and divination).

Bennett (1998) also discussed the divine nature of randomness in some detail. In Homer's *Iliad*, Hector shakes lots for intervention. The Roman historian Tacitus also recorded a ritual used by Teutons to cast lots. In the Hebrew Bible, lots are frequently used in the division of property from inheritance or conquest. And in defense of Masada against the Romans during the second Temple period, men inside the fort drew lots as executioners of the inhabitants with the last individual committing suicide rather than surrendering. Similarly, in Asia, *I Ching* was considered an oracle, which involved the use of hexagrams.

Bennett (1998) continued demarcating significant events in the history of randomness. She noted that Richard de Fournival was the first to write about equiprobability, dice throws, and set versus sequences during the thirteenth century. Additionally, the first author to detail a mathematical study of games of chance was Girolamo Cardano in 1564. In 1654, letters between Blaise Pascale and Pierre de Fermat opened the era of the formal study of probability. Similarly, Mlodinow (2008) had observed the philosopher Hans Reichenbach (1891-1953) was the first to speculate that individuals who lacked training in statistics and probability would have difficulty identifying random series.

Pattern Recognition

Pi, Liao, Lui, and Lu (2008) defined the theory of cognitive pattern recognition: "an object, which is initiated through a visual stimulation (pattern) from the environment, is matched to memory where it is categorized for storage" (pp. 434-435). The authors noted several current models of human pattern recognition: Theory of Template, Theory of Prototype, and Theory of Feature. Templates were described as pre-existing patterns that have been encountered in the past and stored in long-term memory. Therefore, when physical stimulation occurs, a matching process takes place. The authors noted the limitation of the template theory as it did not offer a solution in dealing with new, unobserved patterns. The notion of a prototype, however, went beyond the pattern-copy method and transformed the stimulation into an abstraction of the pattern and its associated attributes. Therefore, when a physical stimulation occurred, broader recognition of patterns was possible. Similar to prototypes, recognition of the components of a pattern decomposed the object into smaller elements. Finally, the authors described the theory of features related to perceptions of patterns and forms. Pattern matching in this theory occurred by shapes and is commonly used today in developing artificial computer intelligence.

Pi et al. (2008) continued their review with a general discussion of human memory, which included both the classification and storage of historical events or patterns for the recall in the future. Furthermore, they separated memory into the time elements of instantaneous, short-term, and long-term. Employing a memory model, the authors included not only the acquisition of additional memory items through time but also the loss of information.

As previously mentioned, this classification process enabled the ability of humans to create a memory. Pi et al. (2008) provided multiple schemas that detailed the organization of human knowledge itself. Their symbol-net model uses a node approach where unidirectional arrows linked concepts within memory at multiple levels. The level semantic-net model detailed by the authors outlined the multidirectional interlinking of the node-concept in a hierarchy with perceived attributes. And, the activation-diffusion model linked concepts (nouns) with attributes (adjectives) while also allowing knowledge creation through the connections of attributes and stimuli not yet perceived. Therefore, this model includes the possibility of anticipated perception.

Pi et al. (2008) concluded with a framework of pattern recognition which accounted for inputs of pattern collection and followed by patterns, which were influenced by both features-prototypes and knowledge-rules. This process produced a matching judgment, and therefore, pattern recognition.

As an example of sequence pattern perception, Restle (1970) tested how college students learned serial patterns discovering that subjects divided the sequences into smaller parts, which could be generated using multiple mental rules, then reconstructed these back into longer sequences. These mental linkages of short patterns are analogous to phraseology in human speech or movements found in themes of music. Restle (1970) proposed that a perceived series also could be decomposed into a set of structural trees within a hierarchy that cascaded to multiple levels of binary outcomes. Furthermore, in experimental trials of pattern sequence memorization, the author showed that subjects were able to recall strings of up to 32 digit-patterns.

Giakoumakis, Papaconstantinou, and Skordalakis (1987) listed the three traditional approaches to pattern recognition including: statistical, syntactic, and hybrid. The statistical approach involved classification by methods such as discriminant analysis or decision theory. The syntactic method used deconstruction to subparts to create primitive patterns at the most elemental level. Lastly, the hybrid approach comprised a combination of both methods. As Restle (1970) had used both decomposition and rules in his experiments, Giakoumakis et al. (1987) compared a rule-based pattern recognition system to the other three methods. They found that the combination of hybrid systems using a decomposition method was similar. The hybrid approach was considered preferable with perceived attributes but was also considered more laborious.

Wang and Hill (2005) noted that the most challenging types of pattern recognition were both temporal and dynamic. The authors proposed a mathematical recognition approach by representing the dynamic aspect through time-invariant models, the similarity of systems dynamics, and state synchronization. Several models were created, tested, and discussed.

And finally, Kriz (1993) discussed how pattern formation in mind and its association with human thought could be used to understand better and treat mental illness. These patterns in mind translated to exhibited individual behaviors. These behaviors could be dysfunctional and occur as individuals interact through "mutual understanding" where each player fulfilled their role, which was determined by the rules of the game set by familial dynamics. The author also described the Bartlett method where an individual would reproduce a visual pattern, which, in turn, was used as a stimulus for the next reproduction, much akin to the theory of rumor transmission. And, previous research by Stadler and Kruse had shown that random patterns of dots were redrawn by individuals into highly ordered structures (as cited in Kriz, 1993, pp. 164-165).

Expertise and Decision Making

Salas, Rosen, and DiazGranados (2009) examined the types of constructs and studies that used expertise-based decisions through intuition. The authors summarized several studies that showed how experts used instinct to make decisions and encapsulated this literature under the construct of naturalistic decision making. They proposed that expertise, adapted from past successful performance processes within a specific domain, combined with intuition, could create rapid and effective judgments. The authors described the research of two internal systems of human cognition. System I contained unconscious information, reflexive, automatic processes, and system II housed the conscious, reflective, and controlled cognition processing. Therefore, among the many mechanisms of expert performance, pattern recognition occurred through a system I rapid perception of larger and significant patterns based on learned, domain-specific expertise.

Salas et al. (2009) suggested that Biggs and Wild (1985) exemplified this rapid pattern recognition encoding. Biggs and Wild (1985) conducted two separate experiments to examine an accounting auditor's ability to estimate certain income statement information based on historical trends using exponential, linear, or logarithmic changes. However, the authors determined bias towards unaudited estimates, which acted as anchors and produced tendencies to overestimate positive trends or underestimate negative ones. For optimal future assessments, the authors suggested a three-step estimation process, which included acquiring information, recognition of an actual pattern, and creating an accurate estimation. However, this process relates to a system II method and not a system I method as proposed.

Rapid pattern encoding was also studied by Klein (1993), beginning with the examination of protocols used by ground commanders during fire situations. The author developed the recognition primed decision-making (RPD) model, which combined the two mental processes of simulation and situational assessment to accomplish tacit decision outcomes (Klein, 1993, pp. 138-139). Three levels of RPD were suggested. The first echelon was a simple match of situational pattern recognition leading directly to implementation. The second echelon included as an extra step of developing a course of action through the generation of internal mental scenarios. And lastly, the third and most complex echelon included an iterative process for event-scenarios, which were similar but not exact matches to previous encounters. Klein (1993) described this internal iterative process as a repeated simulation to generate workable solutions. Therefore, the features of the RPD model included: a situational assessment which excluded judgments of option superiorities, the incorporation of expertise towards a single-event outcome that eliminated multiple-option assessment, and finally, a selection process that satisfices or was good enough to initiate any action.

Ross, Klein, Thunholm, Schmitt, and Baxter (2004) described how the U.S. military was an early adopter of the RPD model. The military decision-making process (MDMP) had been considered burdensome and slower at implementing action plans due to multilevel steps in outcome evaluations. The authors showed that in early field tests against the traditional operational processes, the pace and speed of scenario implementation was increased by 20%.

Documenting the research tensions between naturalistic decision making and heuristics and biases, Kahneman and Klein (2009) collaborated to outline points of agreement and divergence. Generally, both authors agreed that expertise could lead to superior, intuitive judgment, but also that judgment could also be misapplied to produce biased outcomes.

Furthermore, they advised that the internal source of these internal judgments could lead to questionable choices, even among experts. Therefore, when non-experts would follow similar schemas, these individuals might be even more susceptible to these biases or errors. The authors recommended that experts should understand the limits of their know-how and the boundaries of their applicability.

Furthermore, Kahneman and Klein (2009) recommended that the criteria for obtaining expertise should include an event-environment with the visibility of highly verifiable and positive outcomes. Conversely, in unpredictable environments, individuals should recognize that some successful outcomes could result from random chance. They further developed the concept of "fractionation" of skill or expertise into smaller components which occurs, for example, in health care by doctors, nurses, aides or in accounting by expertise in either accounts payable, tax, or fraud subdomains.

This fractionation of domains or decomposition-recomposition, as discussed previously, is demonstrated by Restle (1970) for complex strings, in the entirety of human pattern recognition, from attributes to actions or in-memory storage as in Pi et al. (2008), and in the traditional methods shown in Giakoumakis et al. (1987). And as will be explained subsequently, human pattern recognition and sequence perception are important in the application of Nelson's rules for determining non-random observations within the SPC chart time series. Traditionally, heuristics and bias have greatly influenced random sequence perception (RSP). However, with the application of certain rules used to identify non-random sequences, the ability to decompose these types of series becomes increasingly routine and algorithmic. This ability to discern differences to randomness will continue to improve with increasingly sophisticated software packages.

Randomness

Cognition of Randomness (COR) and Random Sequence Perception (RSP). For this review, cognition of randomness (COR) is defined as a broader and more general study of randomness affecting many domains. At the same time, random sequence perception (RSP) designates the study of randomness in terms of various sequence constructions such as a succession of coin tosses. Within RSP, series can be either evaluated or created by subjects and are usually studied as a dependent variable. Olivola and Oppenheimer (2008) defined random cognition as “a set of beliefs, mental representations, schemata, and reasoning processes that people use to think about random events and processes” (p. 991). Ellis (2000) suggested entropy as a randomness measure, while Dieguez, Wagner-Egger, and Gauvrit (2015) offered an algorithmic definition of randomness.

Furthermore, COR uses a variety of simulations and tools from television screens to medical imaging equipment to deepen randomness investigations. And it can be employed as either a dependent variable or an independent variable to study the effects of another behavioral variable. However, these distinctions were observed in the literature and not explicitly stated as definitions.

Bar-Hillel and Wagenaar (1991) summarized much of the earliest RSP research between 1953 to 1968 and classified the most interesting studies by author, year, types of symbols and their respective counts, lengths, pace, and medium (i.e., written, oral, or other). In reviewing a sample of previous studies, the authors listed the various measures that were used in experiments, which included diagrams, trigrams and associated frequencies, runs, alternations, and conditional probabilities.

Nickerson (2002) distinguished between random processes and random products by describing the former as a property of a process and the latter as the production output of a process. He noted that random products could be considered an item characteristic and independent from its production source. From prior research, he explained that Lopes had suggested a dichotomy in the idea of randomness because sequences could have deterministic structure or could be adjusted non-randomly to "appear" random (as cited in Nickerson, 2002, p. 331). A similar paradox was exposed when considering that a series containing ten consecutive "heads" coin flips with an extremely low probability of occurrence would reflect the same probability of any other named string *a priori*.

Nickerson (2002) also discussed several attributes of randomness. The first, equiprobability, stated that outcomes were equally likely if each has an equal probability of selection. However, he considered this condition as finite, problematic under continuous or infinite conditions, and difficult in actual execution. The second attribute of randomness was unpredictability, i.e., no memory between events. The third was an irregularity, which described a level of entropy measurement. Moreover, the final was incompressibility, which affirmed that the sequence could only be described by the actual string itself and not through any decomposed components.

According to one school of thought, Nickerson (2002) argued the difficulty of determining whether or not process outputs were random or that the prescribed tests for randomness were aimed at particular varieties of non-randomness. However, in another school, Taleb (2007) defined randomness as simply "incomplete information." And when asked about the differences between randomness and deterministic chaos, he responded that any distinction was nonexistent if randomness was considered unknowledge or opaque knowledge.

Furthermore, Taleb (2007) suggested another possible bias between randomness, determinism, and probabilistic thinking. The author proposed what he called the ludic fallacy: the association of the randomness of life events with probabilistic randomness found in casinos (see also Taleb, 2014, p. 184; Bennet, 2019, p. 1). To illustrate this problem of differentiation, he created two hypothetical individuals, Fat Tony, a gregarious man searching for the next sucker, and Dr. John, a painstakingly reasonable intellectual. To contrast their approaches on a series of fair-coin flips, he portrayed their responses after a series of 99 straight heads. When asked what would be the probability of the 100th flip, Dr. John's prediction was 50-50, while Fat Tony's did not exceed 1% because "the coin gotta be loaded" (Taleb, 2007, pp. 123-124). Similarly, Silver (2012) called Dr. John's viewpoint the illusion of the sucker.

Ayton, Hunt, and Wright (1989) were similarly critical of the way psychology had studied randomness. The authors suggested that using inductive reasoning, *a priori*, did not provide criteria for randomness discernment. And retracing Popper, they noted that the act of observation always included an initial point of view, a determination of problems to study, and interests and viewpoints. One primary criticism which the authors discussed at length was instructional bias created by the researchers themselves. Instructions for the questions in an experiment might contain clues to the definition of randomness or patterns. Examples included subjects being asked to produce sequences like a fair coin, to make judgments of randomness by sets of strings that are more patterned, or to use diverse graphical representations of random walks versus coin-tossing nomenclature.

Furthermore, the authors adamantly advocated that any attempt to assess the randomness of a process was similar to attempting to prove a null hypothesis. And they suggested that repeated assessments would increase the possibility of family-wise error and erroneous results.

To remedy the above pitfalls, the authors recommended requiring an *a priori* assumption of the frequency of occurrence based on theory.

Error Types

When assessing randomness, an evaluation is made based on individual outcomes. While subjects can correctly assess randomness or non-randomness, two other possibilities can occur. Henderson et al. (2008) described two error types in terms of common versus special cause variation. The first occurs when mistaking a common cause for a special cause, which is called a type I error, in the reverse case, mistaking special cause for a common cause creates a type II error.

Similarly, in scientific hypothesis testing, a type I error occurs when rejecting the null hypothesis when it is true, i.e., claiming an effect exists that is, in reality, untrue or a false positive. Conversely, a type II error occurs when not rejecting the null hypothesis when it is false, i.e., missing an effect that is, in reality, true or a false negative (Christensen, Johnson, & Turner, 2007, pp. 422-423). Moreover, the type I error is associated with the level of significance defined as a confidence level or alpha, while the type II error is associated with power or beta. Table 1 below details the typical hypothesis testing scenario.

Table 1.
Hypothesis Testing and Error Types

Hypothesis Testing			
		Null Hypothesis	
		Random	Non-Random
No true effect: Stochastic	Random	Correct	Type I
True effect: Deterministic	Non- Random	Type II	Correct

Note. Error types are oriented upper right for type I and lower left for type II.

In Table 2 below, the error types are shown for determining whether or not a sequence is random (RSP) using an identical format to Table 1. Also, this is the table that will be used to score the SPC chart results.

Table 2.
Random Sequence Perception and Error Types

Scoring Table			
		Individual Perception = Is the Series Random?	
		Random	Non-Random
No true effect: Stochastic	Random	Correct	Type I
True effect: Deterministic	Non-Random	Type II	Correct

Note. Error types are oriented upper right for type I and lower left for Type II.

Within the accounting and auditing literature, much has been written about type I and type II errors. A typical firm audit and compliance assessment will have similar outcomes to hypothesis testing and RSP (Sorensen, 1969, p. 556). Table 3 displays these outcomes.

Table 3.
Auditing Results and Corporate Compliance

Auditing Firm			
		Audit Results	
		Certified	Non-Certified
No true effect:	Compliant	Correct	Type I
True effect:	Non- Compliant	Type II	Correct

Note. Error types are oriented upper right for type I and lower left for type II.

Harris and Osman (2012) also provided a similar four-box table based on the controllability of any event-situation, which also follows the nomenclature of the previous two tables and is seen in Table 4. All three tables reflect possible outcomes of human perception within a time series of finite events, which allows for the assessment of possible correct determination or errors *ex post facto*. While the study of COR has led to the identification of many heuristics and biases, one of the most frequently studied is an illusion of control (IoCON), which was first proposed by Langer (1975). Table 4 displays this error as type I and its counterpart, illusion of chaos (IoCHA), as a type II. These two constructs will be explored later; however, as will be shown in the subsequent section, the illusion of chaos or type II error generally has been less frequently examined.

Table 4.
Control of a Situation and Error Types

Psychology (Mental Health)			
		Act as though the world is…	
		Uncontrollable	Controllable
State of the World	Uncontrollable	*Correct rejection*	Type I: *(Illusion of Control)*
	Controllable	Type II: *(Miss or Illusion of Chaos)*	*Hit*

Note. Adapted from Harris & Osman (2012) p. 36. Error types are oriented upper right for type I and lower left for type II.

The following sections contain examples of type I errors that have been extensively studied within psychology and behavioral economics and are organized by theme, relevance to the business field, frequency of study, the commonness of occurrence by practitioners, and personal interest. However, these inclusions are by no means comprehensive.

Hot Hand and Gambler's Fallacy. A positive recency effect has been named the hot hand fallacy, which was originally studied in basketball shooting by Gilovich et al. (1985) and was based on the belief that NBA players could have non-random streaks of successive hit rates in shot-making. However, initially, those strings were found to be random when comparing successes to a 0.5 random generator (Bar-Hillel & Wagenaar, 1991, p. 437). Paradoxically, the authors noted that successive strings of 11 X's (hits) and 10 O's (misses) were considered random with the respective probabilities of 0.00049 (0.5^{11}) and 0.00098 (0.5^{10}), respectively, which was quite improbable.

Bar-Hillel and Wagenaar (1991) also discussed the opposite effect of positive recency as one of the earliest biases dating to Reichenbach. The gambler's fallacy, also a negative recency effect, occurred when roulette players expected certain changes of color (either red to black or vice versa). Similarly, the authors cited Gold and Hester, who determined that gamblers believed that changing or setting aside a particular coin could reverse a current pattern or series of heads or tails outcomes when experimenting with coin flips.

Ayton and Fischer (2004) contrasted the gambler's fallacy and the hot hand fallacy. For the latter, they proposed that player confidence would be gained from a series of successes to create additional successes. And the authors cited Gilden and Wilson for showing streaks in golf puts, darts, and with detecting signals visually or auditorily; Dorsey-Palmateer and Smith for streaks in ten pin bowlers; Smith for modest streaks in horseshoes; and Adams for momentum in pocket billiards (p. 1370).

For the former, the gambler's fallacy, Ayton and Fischer (2004), performed several experiments. In their first experiment, they asked subjects to predict outcomes based either on a gambling condition (roulette) or forecasting condition (by an algorithm). As expected, the longer series were more likely to see a switch in color (negative recency). Furthermore, as successful predictions increased, subjects' confidence also did (positive recency). In their second experiment, after individuals reviewed strings based on alternation rates between 0.2 and 0.8, subjects were more likely to attribute streaks of lower alternation rates to human skill and higher alternation rates to randomness. Similarly, Poundstone (2014) reported that individuals had an internal expectation of an alternation rate, which coincided with 0.75.

Xu and Harvey (2014) studied the betting habits of 776 on-line gamblers and reviewed 566K bets. To examine the hot hand, the authors reviewed the winning sequences and showed that the probability of winning increased with successive bets. As for the gambler's fallacy, the authors reversed the method and reviewed losing sequences. The opposite effect occurred, i.e., the probability of loss increased with each successive one. When attempting to control for skill level, they also discovered that the acceptable odds to gamblers during a positive streak decreased (safer). In contrast, the odds accepted during losing streaks increased (riskier). The authors suggested this behavior was related to a more complex gambler's fallacy, i.e., based on the perception that the positive or negative streak would eventually reverse itself and bet accordingly.

Stöckl, Huber, Kirchler, and Linder (2015) investigated the hot hand and gambler's fallacy using investment experiments among individuals and groups of investors. For the hot hand fallacy, they examined data for streaks of correct decisions made by individuals who had the option to use expert advice. The authors found as streaking increased, both individuals and groups increased the use of expert share in decision-making. Evidence for the gambler's fallacy came from reviewing streaks of alternation where both individuals and groups selected the opposite direction more often than random.

Scheibehenne, Wilke, and Todd (2011) examined whether the hot hand or gambler's fallacy was a heuristic bias or an adaptive strategy. They suggested that an individual assumption of patterns in series or strings was rational thinking based on patterns in nature, i.e., the availability distributions of animal, plant, and water resources or animal foraging strategies. Their experiments assessed adaptation to both negative autocorrelation (high alternation) and positive autocorrelation (low alternation). For positively autocorrelated strings, the authors found significance in the use of a win-stay/lose-shift (stay-shift) strategy. Conversely, with negative autocorrelation, significantly fewer individuals used a win-shift/lose-stay (shift-stay) strategy. Interestingly, individuals, once trained, improved results with positive autocorrelation, but results worsened with the negatively autocorrelated strings.

Tyszka, Zielondka, Dacey, and Sawicki (2008) asked participants to predict subsequent values in a series of strings and observed four types of strategies: a long-run momentum strategy, long-rung contrarian strategy, a short-run momentum strategy, and a short-run contrarian strategy. These strategies were similar to those sometimes used by stock market investors: to buy/sell in rising/falling markets or by taking the contrarian position. The authors witnessed that as perceived randomness increased, participants employed a contrarian strategy, which contradicted the findings of Ayton and Fischer (2004). Also, Poundstone (2014) noted some positive recency and the hot hand effects in the forecasting of business trends, i.e., the near future would remain as it has been among the most recent yesterdays.

Local representativeness. Representative judgment bias, another type I error, was first discussed by Kahneman and Tversky (1972) and occurred when the individual expectation of a small sample drawn from a large population would have extreme similarity to a ratio in the parent population. That is, smaller samples drawn from a larger population were expected to have the same proportions as large sample draws. This bias is also known as the law of small numbers (Bar Hillel & Wagenaar, 1991, p. 444). Research by Kubovy and Gilden also found that students who coded multiple choice answers based on coin flips, kept the constant ratios over short runs (as cited in Bar Hillel & Wagenaar, 1991, p. 438).

A specific version of representativeness was described by Falk and Konold (1997) based on implicit encoding. In producing sequences of random characters, subjects would develop excessive alternating patterns and had difficulty encoding strings. They suggested that the mental difficulty level was a better predictor of randomness than by any objective assessment. Sanderson (2009) devised a calculation to determine the level of encoding difficulty based on the number of runs and alternating runs and devised an algorithm based on a hidden Markov model of duplication and replication. The author also cited Griffiths and Tenebaum, who studied perceptions using mirror images: where the second half reflects the first, where sequences were reversed, and where the first half of the string was duplicated once (as cited in Sanderson, 2009, p. 5).

Lopes and Oden (1987) tested whether or not subjects could detect randomness between strings generated by a Bernoulli process of $p = 0.5$ or those with alternation or repetition. Detection rates of non-randomness were statistically higher than chance, but even graduate students in statistics provided results that were similar to inexperienced subjects (see also Bar-Hillel & Wagenaar, 1991, p. 436). The authors also concluded that this type of randomness study might not be sufficient for real-life judgments or decisions.

Williams and Griffiths (2013) suggested an explication for poor individual performance in assessing randomness was due to a nested hypothesis, i.e., randomness from systematic processes could be non-nested and favor either a heads/tails outcome or nested which favored neither. In a series of experiments, the authors determined that subjects were able to distinguish the non-nested cases and recommended a search for factors affecting randomness judgment beyond general bias.

Several other common biases towards type I error were offered by Tversky and Kahneman (1974). The following deserve mention as they continue to reoccur regularly: Insensitivity of sample size, misconceptions of regression, retrievability, imaginability, illusory correlation, and anchoring. Insensitivity to sample size occurs when statistics from a smaller sample size create the perception (illusion) of much larger variation than what occurs in larger sample sizes. The misconception of regression occurs when regression towards the mean is not conceived as an explanation for more extreme value, i.e., a large variation event may be offset by a subsequently smaller one. Retrievability occurs when the frequency of something is constructed by readily available instances and not a probability, and imaginability describes a flawed assessment outcome that dictates the course of action based on extremely improbable events. The illusory correlation arises when associations are made by suggestion, ease of memory retrieval, or mental construction. And finally, anchoring occurs when an initial estimate is presented, and approximations are made around that value regardless of its accuracy.

Innate random generators. Bar-Hillel and Wagenaar (1991) reviewed whether or not a random generator might exist innately within the human mind; however, they quickly cited three reasons against it. First, accurate random generators were difficult to observe in the natural world. Second, if it existed, its functionality in mind would be questionable because of other biased belief systems that individuals might retain. And third, it perpetuates a belief of an internal homunculus.

While this literature review did not uncover any medical studies of randomness and brain scanning methods such as MRI techniques, future studies may incorporate image techniques while subjects are assessing random or non-random strings. Cabeza, Rao, Wagner, Mayer, and Schacter (2001) completed a similar study by scanning brain regions and having subjects assess true or false statements or perform memory tasks.

Randomness and real life. Bar-Hillel and Wagenaar (1991) suggested the importance of judgment between randomness and determinism occurred when such knowledge provided a guide to real-life. They questioned whether alternation or repetition was a more appropriate approach in the real-world. Some researchers, such as Lopes did not incline to speculate either way. Scheibehenne, Wilke, and Todd (2011), as previously noted, believed that foraging strategies or recency effects were ecologically more important in nature. Ayton, Hunt, and

Wright (1989) echoed this sentiment, although Bar-Hillel and Wagenaar (1991) indicated that the utility or payoff of recency might be recognized by humans more than the alternation. While examining repetition in nature, Mandelbrot (1977) has shown the scaling behavior and the power law of a fractal world would also point to recency. Alternatively, Bennett (1998) remarked that children who generally have an ingrained sense of fairness might wonder if the chance was "fair" (p. 13). She noted that this idea of fairness was interwoven with sharing or "taking turns" whose outcomes are largely different from repeated coin flips. From this view, higher non-random alternation would be associated with equitability while chance with equiprobability.

Taleb and Goldstein (2012) discussed real-world randomness from the viewpoint of the type of distribution assumption associated with particular events. The authors argued against the use of the typical random Gaussian distributions as they expected real-word distributions would be characterized by excessive standard deviations, skewness, and kurtosis, which drastically affect outcomes when errors occurred. Moreover, they argued that in cases of reduced information, the actual distribution type might not be known or was perhaps unknowable, and as a hedge, recommended limiting one's exposure to extreme measurement errors.

Moreover, Kimhi and Zysberg (2009) studied randomness, locus of control, and rare life events. They defined randomness as a "chance" event occurring within a series. The authors generated eight vignettes, four related to self and four related to others, and included both positive and negative outcomes. And they examined whether these events were considered accidental/non-accidental or predetermined/not-predetermined. As expected, subjects with a higher external locus of control were more likely to attribute randomness to rare events.

Similarly, Zysberg and Kimhi (2011) again studied randomness and life events but examined the effects based on gain, loss, neutral outcomes. The authors employed a similar questionnaire to the 2009 study and used a dice throwing procedure to determine the gain (highest number), loss (lower number), and a neutral (neither). Interestingly, individuals in the neutral condition perceived outcomes as most random and maintained a firmer belief in a predetermined effect when compared to the other two conditions. However, these results were dependent on controlling for gender and religiosity.

Ebert and Wenger (2011) conducted two experiments to understand whether or not randomness could be ascribed to free will. The authors described the belief paradox among individuals, stated as “free will was incompatible with a fully deterministic universe” (p. 3). Based on the theory of apparent mental causation, the authors postulated that because individuals believed that intentions caused actions, any movement of randomness could be mistaken for free will. In two separate experiments, the authors examined both actions of the individual-self and perception of the actions of others. While they did find evidence for the association of randomness and free will, several qualifications were enumerated: outcomes had smaller effect sizes; with complexity, indeterministic actions could be seen as deterministic; randomness could be mistaken for free will if that action appeared to aid (or was rationalized after the fact as helping) a perceived goal; and lastly, the problem of determining whether or not individuals actually possess free will.

Learning and randomness. Similar to the encoding of Falk and Konold (1997), Bar-Hillel and Wagenaar (1991) also reviewed a study by Neuringer, who attempted to teach individuals how to assemble non-biased strings of a random sequence. Subjects began by generating 60 series of 100 binary sequences, and in subsequent sessions, the feedback was proposed for improvement with each successive attempt. This feedback included alternations, pairing patterns, and runs of different lengths. The author concluded that training was difficult, that subjects could not learn "randomly," and that the retention of the knowledge gained was short-lived (pp. 441-442).

Randomness and memory. Bar-Hillel and Wagenaar (1991) also reviewed historical literature and the role of memory in randomness perception. The authors noted one study by Kubovy and Psotak who asked individuals to choose a digit between zero and nine, and through histograms, observed that the subsequent distribution was not uniform. Similar studies have verified patterned responses for color, furniture, or fruit, and with coin tosses where heads, for example, was selected 80% of the time. Furthermore, the authors observed that short term memory capacity was usually limited to seven items, plus or minus two; therefore, producing long strings without excessive repetition was problematic. And they described how individuals often formed sequences through an imaginary moving window of repetitive strings. Similarly, Poundstone (2014) explained how early mentalists used these more frequently cited responses and similar techniques to fool the public in their demonstrations of telepathy. The author summarized how Goodfellow had debunked this pseudoscience through his Zenith experiments in the late 1930s.

Olivola and Oppenheimer (2008) explored the interactions between random perception and memory of past events. In their first experiment, the authors tested reconstructive memory, i.e. past memories that generated the same biases occurring in current memory. They found that subjects recalled shorter string lengths if they thought the string was random. Conversely, if subjects perceived it as deterministic, i.e., created by an algorithm, no recall bias was found. In a second experiment, the authors tested the effect of placing a long streak within shorter ones and determined that a series was considered more random when the streak occurred in the middle rather than at the beginning or end. And in a third experiment, they examined the perception of randomness based on both the streak and series length. When recalling a string from memory, subjects would equally judge short and long streaks to be random. However, those who were able to view the entire string continuously, i.e. always visible, saw randomness only in the longer strings.

Zhao, Hahn, and Osherson (2014) suggested the possibility that individuals might be able to separate randomness from non-randomness but might not be able to classify the resulting groups into either category. They also proposed that coding the string itself and the ease of its memorization would affect random perceptions. The authors found that subjects were better able to differentiate between strings than to classify them but also discovered that mirror images were less likely to be labeled random.

Illusion of control. Langer (1975) defined an illusion of control (IoCON) as "the expectancy of personal success…inappropriately higher objective probability would warrant" (p. 311). Referring to Table 4, this behavior would be a type I error when individuals perceived more control than exists. Langer (1975) provided six experiments by introducing attributes such as competition, choice, familiarity, and involvement in chance situations, which created excessive confidence. Similarly, Rodin and Langer (1977) examined the effects of choice and responsibility within a nursing home environment and observed that high levels of choice opportunities positively affected subjects based on ratings by staff and other health metrics.

Presson and Benassi (1996) performed a meta-analysis of the construct of IoCON. At that time, which was 20 years after Langer's publication, the authors remarked that its popularity as a subject of the study had increased, and her article had received over 400 citations. A recent Google Scholar search shows a count of about 5,200, which indicates continued exponential growth. Presson and Benassi (1996) reviewed several situational variables. The first, choice, demonstrated IoCON when individuals preferred their purchased lottery tickets to an option of exchange for a non-purchased one. Another example showed that allowing subjects to choose target bets produced additional confidence in dice rolling.

Similarly, when viewing task outcomes, those subjects with early successes thought their control exceeded others, i.e., even when those outcome sequences were demonstrated by a coin toss. Adding personal control in a task study was also shown to increase IoCON , and IoCON increased when subjects preferred to call a specific color before tossing a multicolored cube. Moreover, any active involvement by subjects not influencing outcome was shown to increase IoCON.

Presson and Benassi (1996) also noted a distinction between IoCON as defined above and illusory predictions associated with IoCON. For example, they reported a previous study by Benassi et al. (1981), where subjects perceived more control when they could guess the result of the dice roll before it occurred. But the authors lamented the fact that much of the IoCON research focused on prediction, judgment, capacity, contingency or some unknown variables which were difficult to classify.

Novovic, Kovac, Duric, and Biro (2012) examined positive and negative affect and the illusion of control. Several studies had shown a tendency that IoCON would provide some resistance to depression, and that individuals with higher IoCON were less downcast and despairing. In their experiments, the authors explored IoCON as a trait, by testing subjects in a two-week pre-period before the actual experiment, and as a state, contingent on pre- and post-experimental results. The negative effect was found to have no real impact on IoCON at any time. Positive affect was correlated with IoCON as a trait, and in pre- post-state conditions. Most interestingly, the authors suggested two interpretations of positive affect: either the task led to positive affect which created IoCON or the task created IoCON and perceived success, which led to positive affect.

Fast, Gruenfeld, Sivanathan, and Galinsky (2008) investigated the relationship between IoCON and power in four experiments. Three experiments were designed to determine if power led to IoCON and any effects associated with optimism, self-esteem, or orientation to action. The fourth was designed to assess the interaction of power with a positive mood as opposed to IoCON. To generate the IoCON in the first experiment, the authors reported that when predicting the outcome of a dice roll, who rolled the die (experimenter or subject) made a difference. Those with high IoCON preferred to perform the roll. In experiments two and three, the authors found that high power individuals were more optimistic and showed higher esteem and more action orientation. Only IoCON mediated power, action orientation, and self-esteem. And finally, the mood was determined not to affect the relationship between power and IoCON.

Harris and Osman (2012) challenged much of the research of IoCON by advocating that it was a rational and adaptive behavior. The authors argued from a Bayesian perspective that individuals would have some prior knowledge and, therefore, a prior estimate of probability, which would be updated through subsequent trials. Furthermore, they suggested that the psychological scenarios created for many experiments were not like those situations encountered in the natural world and that those real situations should be more controllable. Beyond Bayes, the authors turned to decision theory for a discussion of controllability by suggesting outputs reflected not only probabilities but also whether the possible outcomes would be either positive or negative. In an example of actions taken after the London Riots, the authors proposed that costs associated with IoCON, a type I error, were less than those associated with IoCHA, a type II error. As previously described, an adaptation of their four-box table is visible in Table 4.

Wagenaar and Keren (1988) designed two experiments in an attempt to distinguish chance from luck. In the first experiment, subjects wrote about experiences that were scaled along twelve dimensions. In the second experiment, researchers created stories based on surprise, consequence, or superstition. The authors confirmed that the perceptions of luck and chance were different. Large benefits were perceived to come through luck, such as winning a lottery. In a specific gambling scenario, luck was applied to the perception of the gambler's betting choice but did not affect the game or its probabilities, per se. Also, individuals perceived that luck could be influenced, which created the IoCON. And when an event was surprising, individuals associated it with a chance.

In a qualitative study, Ohtsuka and Ohtsuka (2010) examined Vietnamese Australian gambling habits. They noted the existence of higher gambling problem tendencies within culturally and linguistically diverse communities, and that a reliable predictor of gambling frequency was a gambler's belief in the control of illusory outcomes. However, the authors distinguished between perceptions of luck between West and East, the former attributing luck as controllable, and the latter, attributing luck as a vacillating flow of plus and minus tendencies. Therefore, understanding of luck's "current" tendency and flow was vital to Australian Vietnamese gamblers. The sample of gambling subjects included a variety of skill levels, and the authors compiled a list of the signs for both good and bad luck and a list of indicators showing if one was in or out of luck. The authors found that common gambling habits, such as the illusion of control, rationalization, and remembering wins more often than losses, were universal traits. Ohtsuka and Ohtsuka (2010) also suggested that cultural beliefs might reinforce some of the illusion tendencies and that superstition might provide a perception of supplemental control. They also observed attribution patterns among subjects where wins were based on internal qualities and losses to externalities. And the authors uniquely observed how random outcomes could be anthropomorphized by suggesting that the game "lets me win." This behavior is similar to how the West treats its stock markets and associated pundit commentary, e.g., "the market reaction."

Similarly, Silver (2012) described a "fish" in gambling circles as usually the worst player at the table who would lose at a much faster rate than those who were winning or breaking even. Specifically, the author observed that a small proportion of bad players supported moderately skilled players. In describing why this occurred, one individual gambler suggested that some players had delusional perceptions about their abilities, which stemmed from a sense of entitlement. For example, when a player would lose to a low probabilistic hand after making a series of optimal probabilistic bets, a "tilting" behavior, i.e., excessively bad play would likely ensue.

Filippin and Crosetto (2015) observed that within economic studies, IoCON behavior had not been supported. Using a self-developed software task requiring the accumulation of individual boxes, each subject received a monetary incentive for each box collected. However, hidden within one box was a "bomb," which would explode if chosen and would result in a total monetary loss. Using four separate conditions, which included an individual choice of boxes and an IoCON condition, the authors failed to find significant differences among subject groups. The authors suggested that IoCON could be displaced through the use of incentives, which had been confirmed in previous economic research. And as Harris and Osman (2012) had discussed, Filippin and Crosetto (2015) also mentioned the hypothetical nature of the scenarios used for some of the psychological tests and suggested a deeper involvement of subjects participating in economic experiments might affect results.

Soyer and Hogarth (2012) reviewed the manner and form of statistical information found in academic studies to determine if interpretations would change based on the method of presentation. In a survey involving 257 economists in academia, the authors tested the interpretation of a dependent variable on certain values of the independent variable and found that respondents exhibited an illusion of predictability, i.e., underestimation of the width of confidence intervals along with the exclusion of the error term in their computations. However, when graphical representations of the error term were presented without the numerical data, estimates made by experts were more accurate, but not when both data and graphs were presented. Neither level of training nor professional rank produced a significant effect. Finally, the authors advocated for the availability of simulations to accompany scholarly articles to enhance understanding and increase the accuracy of inferences.

Interestingly, Taleb (2014), in his discussions with Kahneman, noted that providing random forecasts to individuals increased their risk profiles even if the individuals knew the forecasts were random. Furthermore, when discussing the creation of various forecasting models, Silver (2012) stated that a common error was overfitting, which would reduce accuracy by including randomness in the model, i.e., a type I error. Furthermore, underfitting a model would, in the same way, create the exact opposite effect (see Kolassa [2016] for a similar discussion of error rates in forecasting; see Delsole & Tippett [2016] for forecasting skill).

Illusory perception. Whitson and Galinsky (2008) assessed whether or not the loss of control would affect illusory pattern perception. In six separate experiments using a variety of stimuli, the authors observed that subjects who were likely to see false patterns were associated with the development of superstitions and conspiratorial beliefs. However, these effects were moderated when individuals could contemplate their personal values (see Waggner-Egger & Bangerter [2007] for studies focused on conspiratorial associations; or Lobato et al. [2014]).

Dieguez, Wagner-Egger, and Gauvrit (2015) disputed the findings of Whitson and Galinsky (2008) with the results of three experiments. The authors suggested that a faulty randomness perception had been associated with four types of conspiracy theories through the idea that nothing-happens-by-accident heuristic. However, the authors found no evidence of conspiratorial beliefs and loss of control.

However, more recently, Whitson, Kim, Wang, Menon, and Webster (2018) again reviewed IoCON and its association with conspiratorial perceptions through regulatory focus, i.e. goal pursuit. The authors suggested that regulatory focus could either be promotion oriented when individuals seek opportunities or prevention-oriented when individuals are sensitive to loss. The authors concluded that loss of control in participants with a promotion focus showed higher conspiratorial perceptions. No associations were noted with a prevention orientation.

Most recently, Stojanov and Halberstadt (2020) performed a meta-analysis that included 23 studies and 45 effect sizes. Based on their computations, the relationship between lack of control and conspiracy theories was not statistically significant even while adding several types of moderation. They did suggest that measuring specific conspiracy theories were more likely to provide a significant result than general conspiratorial statements.

Behavioral Accounting and Finance

In this section, some of the behavioral finance literature will be discussed that included subjects who were actual personnel working specifically in the accounting/finance industry, investors in the stock market, or from various university pools where studies used financial topics. Each item mentioned here contained some aspects of finance or accounting in its study design.

As noted previously, Whitson and Galinsky (2008) examined the relationship between illusory perceptions, control, and conspiratorial beliefs. In one experiment, they manipulated their control group by priming a subject for either a stable or volatile stock market. The authors created positive and negative statements about the hypothetical company in a ratio of 2:1, i.e., either 16 positive and eight negative statements or eight positive and four negative statements. The authors found that market volatility priming affected investment choice and that subjects overestimated negative information.

Kahneman and Riepe (1998) offered advice for investors based on investor beliefs and biases. The authors developed several sections offering both rhetorical questions and recommendations to investors and advisors. In the first section, they discussed judgment bias, which accompanied overconfidence, excessive optimism, hindsight bias, and an overreaction to significant change events. The second section reviewed preferential errors such as probability weighting (expected value), perception of changes in value versus changes in state, the value function from prospect theory, the shape of gambles, the use of purchase price as a reference point, framing, risk policies, and short- versus long-term views. And in the third section, the authors discussed living with the consequences of investment decisions, regrets, and the combined effects of regret and risk. The essay concluded with a self-assessment checklist for financial advisors.

Heuer, Merkle, and Weber (2016) investigated the perception of fund managers' skill towards investors using three surveys. The first survey evaluated fund manager skill and investment choice while the second and third assessed skill and volatility among actual investors.

The authors confirmed a tendency of investors to chase returns and invest in underperforming funds. First, they observed that among a large group of possible investments, investors might not correctly assess the random component of returns and ascribe this variation to manager skill. Similarly, the number of available funds did not seem to alter selection behavior. And using a measure of superior returns, called alpha, investors did not seem to realize that volatility affected a fund's ability to generate those high returns. These factors created what the authors termed as an illusion of skill.

Fenton-O'Creevy et al. (2003) examined the illusion of control and trading performance among actual traders. The authors studied 107 individuals from four London investment banks who traded a variety of instruments and made various assessments of risk. Senior managers within those investment banks also had input to the sample of traders selected for the study. To measure the illusion of control, the authors selected a computer task to avoid a long survey and as a novelty to attract participation. Furthermore, trader performance was also determined through management interviews, which examined traders' past successes. The authors found that IoCON was a maladaptive behavior under certain circumstances: it did not provide a performance benefit in either market or risk management; it added less profit to the firm; it did not affect interpersonal performance, and individuals with IoCON were compensated less. To enhance these effects, the authors suggested three courses of action. Firstly, as a learned behavior, training could address the toxic aspects of IoCON. Secondly, management could take steps to change structural or environmental attributes that might induce IoCON behavior. And finally, for individuals with tendencies towards the error, future recruiting assessments of IoCON might enable selection from a better candidate pool if this type of evaluation were available.

Taleb (2004) also noted some trader foibles based on his career in the industry. As Fenton-O'Creevy et al. (2003) had alluded, Taleb (2004) observed that traders overestimated the accuracy of their personal mental constructions (of the market and environment) and did not consider random variation. The author suggested that their loyalty to these mental schemas would also attach them to their current stock positions. Then, if losses were incurred, traders would switch narratives from short-term to become longer-term investors (see Mlodinow [2008 p. 201] for the consequences of the illusion of inevitability). Furthermore, Taleb (2004) suggested that traders did not construct alternative plans in the event of losses and would not readily question their frameworks or evaluation criteria. And in the worst case, traders would deny the reality of actual value as shown by its asset price and would prefer his/her mentally constructed value, i.e., denial and illusory perceptions.

Ackert, Church, and Qi (2016) attempted to understand various factors causing investors to hold inferior portfolios. In four experiments, the authors designed assessments of choice using assets that were 100% negatively correlated. The experiments included a control group that participated in a series of trades with increasing variability in returns while also fluctuating the timing of payouts and the information related to the actual outcomes. In the first two experiments, using a base case and enhanced variability, investors held imbalanced portfolios. With higher variability in returns, the individual bankruptcy rate increased. The last two experiments varied the payout periods, the first by a random selection of the period and the second at the end of all periods. The authors noted that as information available to the investors decreased, balanced portfolios increased. After reviewing all experimental results, they observed that the overconfidence measure showed no significant effect on portfolio balance; however, subjects with less risk tendencies held more balanced portfolios.

Uecker and Kinney (1977) assessed the representative and protectiveness heuristics among practicing CPAs. As previously discussed, the representativeness heuristic, the law of small numbers, occurs when sample characteristics are assumed to be those of the parent population. The protectiveness heuristic is manifested when sampling preference is for the highest dollar value items over smaller ones or when larger sample sizes are preferred to smaller ones, i.e. not employing comparative sample size and error rate tables. The authors constructed a five-case scenario where subjects would select sample size A or B based on invoice and error counts and the calculated error rates. The authors concluded with both positive and negative results. On the positive side, only 9 of 112 consistently selected lower error rates and ignored sample size while 17 of 112 only selected the large sample size. However, on the negative side, 54% made at least one error of representativeness while 37% made at least one error of protectiveness. And 56% made two errors among only five scenarios.

Johnson (1994) tested individuals and groups of Big Six auditors for both memory recognition and recall with time delays. For a fictitious client, simulated working papers were created, which included discussions of risk in product inventory and long-term debt. Subjects were then required to write review comments as a memory encoding step. After a delay, subjects, either individually or after group discussion not requiring consensus, responded to statements related to those papers along with their confidence level on a 9-point Likert scale. The author found that memory accuracy and confidence were higher among groups while also making fewer type II errors. Additionally, he observed that experience did have a positive effect on error recognition but no effect on recall, recognition, or confidence.

McSweeney (2006) wrote a scathing essay suggesting that net present value and its use created an illusion of certainty. In his previous research, he suggested the archetype of an

Idealized Financial Reporter (IFR) who omnisciently recorded every known transaction as it occurred. However, even with this knowledge, he observed that the future would remain uncertain. Moreover, the author suggested that organizations should have processes that generate future activity without illusions of certainty. However, he observed that most cash flow forecasting was based on perceived futures as desired by advocates of certain narratives. And that often, the deification of the NPV technique occurred at the expense of the discussions it could generate.

The audit function and error types. From a chronological perspective, the development of auditing analysis has been extensive and profound. Table 3 showed the potential audit results based on the compliance of the firm. Among this group of studies, several were selected which focused either on the error types themselves or the cost of making these errors. For example, Sorensen (1969) discussed a Bayesian framework for auditing analysis and sampling, which was constructed based on an historical Bayesian prior, the dollar value of potential actions, and the added knowledge of the financial results gained from sampling. With some basic assumptions, the author created a payoff table using the percentage of defectives, costs of inspection or as-is acceptance, and possible losses from which the auditor could revise the prior probability. He concluded by suggesting that this approach offered a supplemental method to enhance auditing science and move away from non-systematic methods.

Also, Kinney (1975) approached the auditing sampling problem from decision-theory. The author defined a type I error as unnecessary audit prolongations, needless changes to statements, or unwarranted qualified opinions. Conversely, he suggested that type II errors would pass unobserved but would encompass the potential impacts of lawsuits and loss of reputation. The author recommended this approach as an additional appraisal technique because of its insensitivity, the estimation of error costs, and as a benchmark to accounting heuristics.

Similarly, Kinney and Salamon (1982) investigated the use of regression analysis and the applied rules in auditing. This study simulated both monthly and yearly results employing the STAR approach used by Deloitte, Haskins, and Sells to detect account events where an audit was advisable (as cited in Kinney & Salamon, 1982, p. 350). The authors also "seeded" type I and type II error events within the 200 audit years. The author concluded that despite an overall type II error risk of 36.5%, the STAR model performed at expected levels.

Lastly, Dopuch, Holthausen, and Leftwich (1987) assessed the possibility to predict types of auditor decisions using historical financial and market variables. The types of auditor opinions reviewed included those that were either clean or first-time qualified along with supplemental qualifications of going-concern, litigation, asset realizing, and those firms with multiple issues. Five financial statements or ratio variables were used, and four stock market variables among publicly traded companies separated by SIC code. A probit model for the assessment was constructed with weighted exogenous sample maximum likelihood procedures. A model type I error occurred when classifying a qualified opinion as clean and vice versa for type II while using measures of relative costs for error types. The authors concluded that the model could predict initially qualified opinions; however, estimated probability scores varied across opinions (see Geiger & Rama [2006] for a going-concern assessment suggesting that the Big Four made significantly less type I/II errors than regional or third-tier firms; and Dunn, Tan, & Venuti [2012] for a similar but more recent study of the Big Six).

There are also a few studies tying error types to corporate fraud, a recurring and costly practice. Among them, Deshmukh, Karim, and Siegel (1998) employed signal detection theory to assess management fraud. The authors constructed the typical auditing four-box decision table, which included "hit rates" as observing signal, "correct identification" as noise, "false alarms" as type I errors, and "miss rates" as type II errors. Likelihood ratios were calculated as hit rates as a percentage of false alarm rates, which were subsequently compared to a criterion value based on prior probability and consequences. The authors suggested that costs of type I errors such as additional working hours, unnecessary adjustments, and lost business and revenue were not evaluated and problematic. For type II errors, they noted that a 10 to 99 multiplier of type I error cost was typical.

Furthermore, they cited a statement made by Arthur Anderson, which estimated that 12% of all Big Six revenue was related to type II error cost (as cited in Deshmukh, Karim, & Siegel, 1998, p. 132). The authors recommended an increased tolerance for type I errors and false alarms as a remedy to combat fraud. They also postulated that this tolerance for error would need to be increased as the disparity between costs of type I and type II errors continued to mount.

Studies of Randomness in Diverse Domains

The previous section has illustrated predominantly type I errors. Understanding how randomness and corresponding error types are viewed and studied in other domains can provide additional perspective, insight, and opportunities to avoid errors where interactions within different fields may occur.

CEO effect. The CEO effect attempts to explain the benefit of a CEO using a variance decomposition method, which dates to Liberson and O'Conner (as cited in Fitza, 2014, p. 1840). In this technique, return on assets (ROA) was separated into parts based on year, industry, company, and CEO. Fitza (2014) suggested an improvement to the variance decomposition method by demonstrating the effect of adding a randomly generated element that would simulate possible performances by "chance." He indicated that performance outcomes could be randomly influenced by both factors within and outside of the CEO's control. To account for this effect and to validate this approach, the author created a dummy variable to replace the ROA using the identical mean and standard deviation of the calculated ROA followed by a 100-trial simulation. Previous studies had estimated the CEO effect between 17% and 19%. The random variable simulations yielded about 14% of the variation, which significantly reduced the CEO effect.

Quigley and Graffin (2017) took exception to the analysis by providing a rebuttal based on the incremental R^2 values and proposed a new quantitative tool of multilevel modeling (MLM), which returned the CEO effect to slightly lower historical levels. However, Fitza (2017) applied his randomization method to MLM and observed that the CEO effect was largely due to chance. All authors concluded that the most appropriate next step was to determine situations when or under what conditions the CEO effect was most important.

Similarly, Parnell and Dent (2009) reviewed firm performance and the notion of luck. They suggested that as management knowledge of firm resources and capabilities increased, management perception of the role of luck increased. And in another study, Kim, Eberhart, and Armanios (2016) employed bootstrap simulations to quantify luck and performance. They determined that 95% of performance outcomes between high and average performers could be attributed to random effects.

Type I or Type II Errors in Diverse Domains

Sensitivities to an error in randomness perception have been widely studied in many domains. In discussing calibration and forecasting by the Weather Channel, Silver (2012) noted a study by Bickel and Kim that showed the Weather Channel overpredicted rain chances. In discussing the reason for this strategy with personnel at the television station, they explained that viewers were more sensitive to failure to predict rain, a type II error, as opposed to false alarms when rain was predicted and did not materialize, a type I error. So, forecasts were expressly increased to avoid failure in prediction. Silver (2012) also interviewed Murray Campbell, who led the team of IBM programmers for Deep Blue before the chess match with Gary Kasparov. The author described a unique type I error committed by Kasparov during the first match of 1997. A computer bug occurred in its 44th move that was completely random. The IBM team had known about the bug and had presumably fixed it. However, during the second game and after a serious blunder by Kasparov, Campbell recalled that Kasparov had attributed that first game random move as a sign of superior computer intelligence and not as a simple error.

Similarly, Poundstone (2014) recounted the development of the earliest "outguessing" machine created in Bell Laboratories of the 1950s. The machine's logic was based on the assumption that individuals would not make random choices. The programmers determined that the best strategy for the machine would be to make a prediction of human behavior when it was winning and play randomly when losing.

Anderegg, Callaway, Boykoff, Yohe, and Root (2014) discussed type I and II errors in the study of climate science. The authors suggested that to lower the possibility of a type I error to avoid political and other "drama," hurdle rates were excessively restrictive. The authors suggested that researchers should equally account for both types of errors and outlined the risks of a type II error: limited actions with the passage of time and precluding communication and debate about the possible risk to human life and world economies. The authors suggested the classification of uncertainty into four types: risk, with known odds; uncertainty with known parameters but without knowledge of odds; ignorance or unrecognized risk; and indeterminacy, which was defined as conditional knowledge intermingled with influence from science, society, and politics. To alleviate this asymmetry of errors, the authors recommended a balanced treatment of error types, the consideration of a wide variety of possible outcomes while drawing information for multiple sources, and the use of elicitation analysis to understand broader impacts.

Davidson (1986) noted type I and type II error asymmetry in the publication of clinical trials. After studying 107 trials, the author found that 71% of those favored new therapies, of which 43% were sponsored by pharmaceutical companies. And of the 31 trials that favored traditional therapy, only 4 had pharmaceutical sponsorship. The author determined a statistical significance between the source of funding and study outcome. However, he also enumerated several reasons for this publication bias. Typically these studies could have used smaller sample sizes which caused weak power calculations and possible type I errors; discontinuation of studies which were showing no effects (see O'Brien & Fleming [1979] for early termination of medical trials; and Freiman, Chalmers, Smith, & Kuebler [1978] for type II errors in negative trials); the possibility that unsuccessful trials would decrease research funding, and the conflicting constituencies of researchers, pharmaceuticals, regulators, and the public.

Similarly, Ioannidis (2005) posed similar questions of the validity and replicability of research findings. The author created a metric called the positive predictive value (PPV) using type I and II error rates, a bias factor, and a pre-study probability of a true relationship (see Silver [2012, pp. 242-255] for a discussion of a Bayesian prior, conditional probability, and frequentist schools; or Mlodinow [2008, pp. 171-172] for the development of significance testing by Fisher). Ioannidis (2005) made several suggestions for improvement in research studies which included increasing power through sample size, an emphasis of the totality of research findings within trials and a de-emphasis of single findings of significance; and a reduction of the prominence of statistical significance while obtaining a better *a priori* understanding of probabilities before attempting a research project. Beyond those recommendations, Forstmeier, Wagenmakers, and Parker (2017) complemented and added to that list with their table of problems and proposed solutions. Several issues that the authors identified included novelty seeking, HARKing or hypothesizing after the results were known, discarding unsuccessful experiments, confirmation and hindsight bias, incorrect interpretation of regression towards the mean, and overfitting of models, among others.

Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, and Woloshin (2007) stated flatly that many patients, physicians, and other individuals did not understand health statistics. The authors suggested that statistical illiteracy was rampant in many fields and illustrated this through several recent examples: the contraceptive pill scare of the U.K. in the 1990s which led to a significant increase in abortions; the inability of gynecologists to understand positive mammograms; misinterpretations of survival and mortality rates; or overdiagnosis, defined as finding a clinical disease which was non-progressive and did not require treatment. The authors also noted that overdiagnosis (a type I error) could have harmful effects when unnecessary treatments were prescribed.

In previous research, Gigerenzer et al. (2007) observed the lack of medical literacy was compounded with the illusion of certainty, which they defined as a need for certainty where none exists. Several consequences of statistical illiteracy were listed. First, patients were susceptible to manipulation, which could lead to additional anxiety or false hope through questionable advertising. Or in the cases of patients with a pseudo-disease, treatments would cause actual harm. Furthermore, the authors remarked that statistical illiteracy would affect the relationship of informed consent and the shared decision making between doctor and patient. And finally, they offered numerous causes and remedies to enhance the statistical knowledge for patients, doctors, and population at large.

Silver (2012) discussed this similar problem of medical illiteracy and extended the discussion to the 9-11 terrorist attack in New York City. He recommended that any assessment include a Bayesian approach based on probability of occurrence, non-occurrence, and a prior probability. And Bennet (1998) began her entire discussion of randomness with the interpretation of a hypothetical tuberculosis test, conditional probabilities, error rates, and associated misjudgments.

Taleb (2014) discussed the “do you have evidence” fallacy and its relationship to iatrogenic, i.e., the study of harm done by the medical personnel. The fallacy expressed the idea that “evidence of no harm” was not equivalent to “no evidence of harm” and was similar to the conclusion in medical diagnostics between “no evidence of disease” and “evidence of no disease” (p. 337). The author further generalized this argument to the confusion generated from “absence of evidence” and “evidence for absence” which occurs when proponents for something argue the former and imply the latter. He also observed that individuals were more likely to “fit beliefs to actions rather than fit their actions to belief,” which could create illusory beliefs or self-deception (p. 415).

As in hypothesis testing, researchers have preferred to minimize type I errors by restricting hurdle rates while increasing the possibility of type II errors. Rizzolli (2016) described these errors within the field of law and adjudication as a choice between convicting innocent defendants, a type I error, juxtaposed to allowing guilty defendants to go free, a type II error. The author noted that type II errors in this field had been extensively studied. However, he observed that type I errors were being examined more frequently with the advent of sophisticated DNA analyses. He estimated a judicial type I rate of 3.5% to 5% between 1982 and 1989.

Similarly, law enforcement faces the same dilemma regarding arrests. Garoupa (1997) examined the idea that private law enforcement would increase the non-offender arrest rate or a type I error as arrests would enhance revenue. The author concluded that without regulation of these private organizations, more crime and arrests of non-offenders would occur. However, if penalties were established for excessive detection (minor crimes) and if incentives were established for under-detection (major crimes), the error problems would be resolved.

Consequences and Costs of Type I and Type II Errors

The costs and consequences of type I or type II errors can be high as was demonstrated firm going concern issues within the domain of corporate compliance. From several diverse business research sources, additional impacts of these errors will be discussed here: from the waste of time and skills defined in basic quality literature, an issue of efficiency, to consequences that could result in financial distress or bankruptcy of the firm, a concern of effectiveness. At a minimum, error avoidance could save resources or lead to opportunities for success. This section details some of these potential influences.

Skhmot (2017) listed the eight types of waste identified in quality improvement and lean quality methodology. The author noted that the first seven were initially included as part of the Toyoda Production System through the TIMWOOD acronym during the 1970s: transportation, inventory, motion, waiting, overproduction, over-processing, and defects. The eighth was added in the 1990s, which was skill waste due to nonuse, misuse of employee skills, knowledge, and talent. The effort spent in reviewing randomness within a process to attribute a false cause or type I error is the same as discarding cash or wasting any other company asset. This type of waste is the inefficient use of personnel. For classifications of waste, see also leanway.net, leansmarts.com, leanmanufacturingtools.org, shumla.com, and kasama.us.

From an anthropological perspective, Graeber (2018) has documented the proliferation of meaningless work, useless jobs, or perhaps brought them to the forefront of discussion. The five job types that he identified included flunkies, goons, duct tapers, box tickers, and taskmasters. The author relayed the frustration of those individuals who lamented specific parts or all of their work and suggested industrialized society has developed a cult of work itself regardless of its benefits. And the author discussed many possible causes, origins, and potential remedies, some of which could be considered controversial. Once again, failures to address special causes or searching for one-off errors within random variation would demoralize any employee. And the manager who assigns these types of tasks would be considered a type of taskmaster.

Harris and Osman (2012) offered an example where the illusion of control might be preferable to the illusion of chaos. As an example, they reviewed the U.K.'s government response to the London riots in 2011. The authors suggested that allowing the riot to continue in an incontrollable state would cost more than perceiving the situation as controllable (illusion of control) even without knowledge of the actual state of the situation. Therefore, the authors suggested that this reaction of assuming an illusion of control state was perfectly rational and reasoned. This schema also suggests that an assumption of IoCON is preferable and least costly when actual knowledge is scarce.

Other literature has suggested the consequences of these errors. Tellis (2013) discussed type I and type II errors concerning the innovation of the firm and undertaking new capital or market expansion projects. He described the type I error as a failed innovation whose costs were associated with research, development, testing, and eventual commercialization. Similarly, he noted that a type II error occurred when a failure to approve a potentially successful innovation. The author described a four-box matrix with the go/no-go tradeoffs and error types.

In some cases, the firm would successfully accept good projects and reject potential failures. Depending on the screening criteria, the types of potential error would change. Looser project hurdle rates lowered type II errors by allowing more ventures but created the possibility of type I errors by approving unsuccessful opportunities. Conversely, tightening criteria decreased the possibility of a type I error as more projects are rejected while increasing the chance of type II errors and missing an opportunity. This argument is similar to the tradeoff of type I and type II errors used in statistical hypothesis testing and quality improvement (see Deming, 1982).

Furthermore, Tellis (2013) stated that the costs associated with type I errors were limited and finite. However, costs associated with type II errors could be large: on the upside, high market capitalization, market impact, and wealth; on the downside financial distress, reorganization, or bankruptcy. The author suggested that type I errors, their measurable impacts, and recency effect might cause management to fixate on these types of error. However, he also observed that the effect of a type II error might not be immediately manifested, but when known, it might be passed the time to react effectively. Therefore, the author encouraged companies to consider both types of error with their innovation decisions, which could be considered as options of smaller and varied opportunities with large but rare possible payouts (p. 73).

Tellis (2013) made two assumptions in his analysis. First, that all projects would be known, and the second, an optimal hurdle rate would be selected. In many business environments, all projects will not be known or sought, hurdle rates may be deficient, and resources are constrained. Therefore, in an economic sense, an opportunity cost is created from those wasted resources which could be deployed elsewhere. Individuals whose selection criteria are biased are reviewing randomness to assign a cause, a type I error, and use resources inefficiently, which should be deployed in exploring actual special cause variation. Furthermore, this behavior might simultaneously increase the opportunity to make a type II error through resource exhaustion, which suggests both types of error might co-occur.

Furthermore, to encapsulate the behavioral impact of IoCON/IoCHA, an obligatory summary follows. These maladaptive behaviors affect not only staff personnel but also management itself. With IoCON, individuals are overconfident, make bold statements, sometimes without follow-up, and create an illusion of certainty where there is none. They can be excessively optimistic, exaggerate their talents while thinking they have more control of their destiny than in actuality. Moreover, they tend towards hindsight bias, "I knew it all along," which increases faulty rationalizations, ignoring or discounting lessons learned, and repeating similar mistakes. And they are less likely to predict negative outcomes or chance events.

Conversely, the effects of IoCHA for the firm and managers could also be serious. Contraction of the number of projects through spurious quantitative criteria and bias for action may well direct the firm towards cosmetic and superficial changes. The creation of analysis paralysis and fixation of type I errors might precede stagnation, financial distress, or worse.

Literature of Operational Terms

Measuring randomness: Statistical process control charts and Nelson's rules. In his seminal work, Shewhart (1931) outlined three postulates for the scientific basis of control in manufacturing: all causes are not alike, chance causes exist in nature, and assignable causes can be discovered and removed. He divided variation into two distinct types. The first was the special or assignable cause, which equated to influences outside the current process (phenomenon) being studied, i.e., non-random error. The second was a common cause, which was created by factors internal to the process, i.e., random. The author completed a systematic review of the history of various distributions using both graphical and analytical examples, specifying their general and particular characteristics: from Simpson and LaGrange in 1756-1773 to Tschuprow in 1910-1916. Shewhart (1931) devoted a large part of his book to explaining the detection of deterministic observations and how to calculate process limits, which would be used in SPC charts. He concluded by stating that the purpose of the quality report was separating common causes from special causes and explicating the actions necessary to eliminate special cause variation.

Wheeler and Chambers (1992) specified how to create the SPC chart used today using a manual method. The chart itself is a graphical representation of observations within a time-series that are plotted along a centerline where each event varies around it. The basic center-line calculation uses the mean or average of the data points and a measure of dispersion, the range, and standard deviation. Control limits, both the upper and lower bounds (UCL, LCL), are numerical values demarcating the minimum and maximum limits that separate common and special variation, the former falling within three standard deviations of the mean while the latter falls outside those limits (see also Mohammed, Cheng, Rouse, & Marshall, 2001, p. 463).

Wheeler and Chambers (1992) also noted four myths related to the use of SPC charts. The first was that data must be normally distributed. However, as noted previously, Shewhart's original analysis discussed many types of distributions. Second, SPC charts successfully function because of the central limit theorem or, the third, that data with serial correlations were inappropriate. The authors suggested that both of these ideas were artificial barriers to chart use.

They indicated that the purpose of the control chart was to obtain insight and not calculate the probability. And finally, the fourth myth stated that observations must be within controllable limits before charts can be used. This requirement is unnecessary since one of their functions is to work towards process control, which implies “out of control” processes. Similarly, Deming (1982) designed the “red bead” experiment, which demonstrated the impossibility of handpicking special cause variation from a larger random sample. In sum, the essence of quality process improvement is to separate special from common cause, eliminate special cause, and further reduce common variation by process changes (see Deming, 1982, p. 321; and Wheeler & Chambers, 1992, p. 6; or Keys & Reding, 1992, for accounting and TQM).

The rules and interpretation of control charts have changed since Shewhart’s original work. As previously noted, his demarcation between common and special cause variation was plus or minus three standard deviations. Noskievičová (2013) provided both a brief history and summary of chart rules (also known as Nelson’s rules) over the history of quality improvement. The table below represents a summary of the historical rules, year developed, and changes over time. Sigma XL and SPC Excel are software packages designed for quality improvement, but many other software packages also include quality modules.

Table 5.
SPC chart rules through time

Sigma XL Rule Definitions	**Shewhart (1931)**	**Western Electric (1958); Boeing, GE**	**Nelson (1984) & ISO 2589 (1991)**	**SPC Excel**	**Griffiths et al. (2010)**	**Noskievičová (2013)**
Test 1: 1 point more than 3 StDev from CL	**Test 1:** 1	**Test 1:** 1	**Test 1:** 1	**Test 1:** 1	**Test 1:** 1	**Test 1:** 1
Test 2: 7 points in a row on same side of CL			**Test 2:** 9	**Test 4:** 7	**Test 2:** 9	**Test 4:** 8
Test 3: 7 points in a row all increasing or all decreasing			**Test 3:** 6	**Test 5:** 7	**Test 3:** 6	**Test 5:** 6
Test 4: 14 points in a row alternating up and down			**Test 4:** 14	**Test 8:** 14	**Test 4:** 14	**Test 7:** 14
Test 5: 2 out of 3 points more than 2 StDev from CL (same side)		**Test 2:** 2 of 3	**Test 5:** 2 of 3	**Test 2:** 2 of 3	**Test 5:** 2 of 3	**Test 2:** 2 of 3
Test 6: 4 out of 5 points more than 1 StDev from CL (same side)		**Test 3:** 4 of 5	**Test 6:** 4 of 5	**Test 3:** 4 of 5	**Test 6:** 4 of 5	**Test 3:** 4 of 5
Test 7: 14 points in a row within 1 StDev from CL (either side)			**Test 7:** 15	**Test 7:** 15	**Test 7:** 15	**Test 6:** 15
Test 8: 8 points in a row more than 1 StDev from CL (either side)		**Test 4:** 8	**Test 8:** 8	**Test 6:** 8	**Test 8:** 8	**Test 8:** 8

Note. (modified Noskievičová, 2013, p. 3) showing different rule number conventions and default values. From right to left: SigmaXL rule defaults, Shewhart's original work, Western Electric, Nelson, SPC Excel defaults, and suggestions from Griffiths and Noskievičová.

The first column in Table 5 lists the rules found in the SigmaXL package, the software used to construct the measures for this test. Each rule is numbered from one to eight; however, each rule number varies over time and by software package. Similarly, each rule contains a quantity of a certain number of observations, which demarcates the division between special and common cause variation. These are also visible in the table. In his original work, Shewhart (1931) only described the three-standard deviation rule. By the late 1950s, Western Electric had expanded their rule list to four. With the quality of revolution that resulted from the success of Japanese companies in the 1970s and 1980s, Nelson (1984) expanded the rule list to eight items while also providing a graphical representation of specific rules.

Furthermore, Nelson's rule two stated nine points in a row on the same side of the centerline. The two software packages use seven points, and Noskievičová (2013) used eight, which correspond to different probabilities of occurrence. However, the default values in most software packages can be adjusted as necessary to the desired probability. While a general standard of the exact number of points does not seem to exist, Griffiths, Bunder, Gulati, and Onizawa (2010) calculated probabilities for a selection of rules and proposed a fixed probability of occurrence with a convergence of ratios of 0.003. The authors also noted that the rules were not mutually exclusive, and that one observation could be assigned multiple rule violations.

Moreover, they computed the probability of rule four, which described sequences of alternation and reviewed early alternation studies dating to Andre (1879, 1881, 1883). Given the tendency, previously noted, of non-random alternation rates, Griffiths et al. (2010) also created a table of probabilities based on alternation rates between three events (probability of 0.667) to 14 events (probability of 0.005). The probability calculation was based on the equation $p = 2.5592e^{-0.452x}$, where p is the probability, and x is the number of alterations (Griffiths et al., 2010, p. 5).

Additionally, Griffiths et al. (2010) calculated the probabilities of each of Nelson's eight rules. Based on the Sigma XL rule differences as noted in Table 5, these probabilities were adjusted, where appropriate, and are visible below in Table 6. Because of the law of large numbers, simulated random occurrences should approach the calculated probabilities in large samples. Table 6 also contains sample sizes of 5,000 and 100,000 trials that resemble actual probabilities of occurrence by chance.

Table 6.
Probabilities of Nelson's Rules and Simulation Results

Tests for Special Cause Variation	Probability of Occurrence Griffiths et al. (2010) adjusted	500 number strings MS Excel [Rand()] for 10 Trials (misconception of chance)	500 number strings Random.org for 10 Trials	10K number strings MS Excel [NORM.INV(RAND(), mean, standard dev)] for 10 Trials	Random10K (Random.org) for 10 Trials
Test 1	0.00270		0.00220	0.00252	0.00257
Test 2	0.01563	0.01400	0.01560	0.01559	0.01544
Test 3	0.00040		0.00040	0.00004	0.00010
Test 4	0.00457	0.00500	0.00180	0.00250	0.00204
Test 5	0.00306		0.00260	0.00161	0.00183
Test 6	0.00553	0.00680	0.00540	0.00440	0.00443
Test 7	0.00478		0.00140	0.00446	0.00507
Test 8	0.00010	0.00020		0.00012	0.00011
Total		0.02540	0.02800	0.03019	0.03059

Note. From left to right, calculated probabilities of individual Nelson's rules, associated simulations of 5,000 and 100,000 trials using MS Excel or random.org.

The uses of either the probability of rule occurrence or a simulation of error frequency are two points of reference for any time series. Using a method of error frequency, one can compare time series of any type with any other, e.g., a sales variance process, a service accounting metric, or a stock price. If the calculated non-random (special cause) variation exceeds random chance, then special cause variation exists in that string. Table 7 below details a large frequency simulation, a six-month sales variance process assessment, Mondelez and SP500 prices for 275 trading days, and a service recovery two-year assessment. Each series contains its associated rate of non-random variation by rule. Also, beside each rule of the series, *p-values* have been calculated to show significant differences from what would occur randomly in a simulation. And using a simulation method also allows for a calculation of overall error rate assessment. The associated *p-values* show whether the series error rate is significantly different from the simulation. For example, in the random trials, 3% error rates are expected by chance versus 13% of sales process, 9% of Mondelez, 7% of SP500, and 4% of service recovery.

Table 7.

Time Series Comparisons Using Nelson's Rules and Error Counts

Tests for Special Cause Variation	**Random10K (Random.org) for 10 Trials**	**6 Month 2019 Sales Variance Sample (n=3,896)**	***P-value***	**MDLZ Daily Returns 29 June 2016 to 30 August 2017 (n=275)**	***P-value***	**SP500 Daily Returns 29 June 2016 to 30 August 2017 (n=275)**	***P-value***	**Service Recovery 2016-2017 (n=24)**	***P-value***
Test 1	0.00257	0.02721	0.000	0.03636	0.003	0.01818	0.053	0.04167	0.338
Test 2	0.01544	0.04723	0.122			0.01091	0.470		
Test 3	0.00010								
Test 4	0.00204	0.00282	0.390						
Test 5	0.00183	0.00334	0.002	0.01091	0.147	0.00727	0.288		
Test 6	0.00443	0.00026	0.000						
Test 7	0.00507	0.05364	0.000	0.05091	0.001	0.03273	0.010		
Test 8	0.00011								
Total	0.03059	0.12782	0.000	0.09091	0.001	0.06545	0.019	0.04167	0.786

Note. Error rates based on Nelson's rules from left to right: random occurrence, 6-month sales variance, Mondelez, SP500, and a 2-year service recovery with associated *p-values*.

An additional method to examine, compare, or benchmark time series processes is through a ratio of special versus random variation. This calculation separates the total variation of a time series into its components by calculating the underlying amounts related to special cause variation and whose difference to the total variation in the series would be the random variation. Table 8 details the similarities of non-random variation in a sales process over a 10-year period is a stable and scalable prediction. The R^2 of this decade of observations was 96% given by the power graph line of $y=0.128x^{1.0437}$. While Sigma XL does not provide these results in the current version, all the elements are present to allow a computation. For the two stocks previously mentioned, the special cause variation for Mondelez and SP500 were 20% and 8%, respectively. Moreover, the totality of this information gives credence to the 85/15 rule attributed to Juran, i.e., 85% of the root causes of errors are attributable to "systems, processes, and structure while 15% can be traced to people" (Clemmer, 1992, p. 67).

Table 8.
Sales Variance: Separation of Random and Non-random Variation

Year	Oversold (Positive Variation)	Undersold (Negative Variation)	Total Variation	Non-Random Variation	Random Variation	% of Non-Random Variation	% of Random Variation
2009	$2,840,207	$5,508,677	$8,348,883	$1,874,309	$6,474,574	22.4%	77.6%
2010	$3,260,460	$5,795,934	$9,056,394	$2,335,479	$6,720,916	25.8%	74.2%
2011	$3,932,765	$5,540,532	$9,473,297	$2,679,310	$6,793,987	28.3%	71.7%
2012	$3,037,403	$5,221,363	$8,258,766	$2,396,269	$5,862,497	29.0%	71.0%
2013	$2,663,584	$8,553,638	$11,217,222	$3,474,567	$7,742,654	31.0%	69.0%
2014	$1,743,874	$7,872,060	$9,615,935	$2,182,877	$7,433,058	22.7%	77.3%
2015	$1,786,043	$8,335,417	$10,121,460	$2,747,146	$7,374,314	27.1%	72.9%
2016	$2,533,672	$4,577,125	$7,110,797	$1,314,652	$5,796,145	18.5%	81.5%
2017	$2,021,944	$1,979,637	$4,001,581	$1,023,995	$2,977,587	25.6%	74.4%
2018	$1,688,443	$1,915,520	$3,603,963	$1,008,100	$2,595,863	28.0%	72.0%
2019 6-monthYTD	$583,204	$542,457	$1,125,661	$261,947	$863,714	23.3%	76.7%

Note. 10-year sales variation, which separates over/undersold and assigns associated non-random variation based on Nelson's rules.

Furthermore, Noskievičová (2013) and SPC Excel (n.d.) provide a list of possible special causes listed by rule based on previous research in manufacturing environments. An adapted version of their tables appears below in Table 9 and includes some unpublished identified causes from an amalgamation of individual sales variances in a construction company.

While most individuals with general knowledge of SPC charts consider their use, specifically in manufacturing or sales environments (see Selden, 1996), the methodology is also applicable to service industries. Henderson et al. (2008) review patient stroke care of 2,962 patients in three U.K. hospitals by retroactively plotting SPC charts to assess common or special cause variation in four areas: brain imaging, prescribing aspirin after stroke, the proportion of patients receiving strong unit care, and the proportion of patients discharged on a statin. Findings included the confirmation of improvements in patient care and those improvements that were expected but did not occur. Moreover, the authors stated that several unexpected signals of special cases were investigated. Some specific causes were determined while others were not.

Table 9.
SPC Chart Rules and Possible Causes

Tests for Special Cause Variation	Type of Unnatural Pettern	Pattern Description	Possible Causes - Manufacturing	Sales Variance
Test 1: 1 point more than 3 StDev from CL	Large shift, strays, freaks	Sudden high change	New person doing the job, incorrect set-up, measurement error, process step skipped or not completed, power failure, equipment breakdown, damaged in handling, gage jumped setting, tool breakage	Included unnecessary items (error of commission), excluded necessary items (error of commission), missing documentation for HVAC price match, large CCA, commission adjustments
Test 2: 7 points in a row on same side of CL	Stratification	Small differences between values in a long run, absence of points near the control limits	see 6	Policy decision (e.g. not couting in sales subtotal), SPC fear of miss-specification, SPC taking advantage of policy variance, SPC using sales tolerance, missing actual units,
Test 3: 7 points in a row all increasing or all decreasing	Mixture	Sawtooth effect, absenced of ponts near the central line	Tooling wear, temperature effects (cooling, heating), operator fatigue, inadequate maintenance	
Test 4: 14 points in a row alternating up and down	Cycle	Recurring periodic movement	Overcontrol, tampering by operator, alternating raw materials, incorrect sub-groupings or sub-groupings from different sources, alternating suppliers, other alternating causes	
Test 5: 2 out of 3 points more than 2 StDev from CL (same side)	Smaller sustained shift	Sustained smaller change	see 1	see 1 with smaller effects
Test 6: 4 out of 5 points more than 1 StDev from CL (same side)	Trends	A continuous change in one direction	Raw material change, change in work instruction, different measurement, device/calibration different shift, person gains greater skills in doing the job, change in maintenance program, change in setup procedure, new operator, failure to recalculate limits with process improvement, fixture change, change in operator motivation	
Test 7: 14 points in a row within 1 StDev from CL (either side)	Systematic Variation	Regular alternation of high and low values	see 8	SPC judgment of type of customer (e.g. sit of prime/plus), winning sale shifts mentality (compensating behavior), small part substitutions (e.g. thermostat, sell one brand substitute another)
Test 8: 8 points in a row more than 1 StDev from CL (either side)			More than one process present (e.g. shifts, machines, raw material), incorrect data or control limit calculation (e.g. after process change)	

(modified Noskievičová, 2013, p. 4; SPC Excel, n.d., pp. 8-9)

Mohammed et al. (2001) provided several examples of retrofitting SPC charts to determine both special and common cause in a hospital setting: mortality rates of children younger than one year from the UK Cardiac Surgical Register, mortality rates from the medical doctor and serial killer Harold Shipman, IVF treatment, prevalence of coronary heart disease among general practitioners, and neonatal deaths. The authors also cited several U.S. studies, which included the study of mortality rates in hospital trauma cases, infection control, and monitoring and detecting outliers in public health reporting. The variety of type and domains analyzed points to the overall applicability of the tool and its ability to gain insight into diverse problems.

Measuring randomness: weakness of SPC charts. Taleb (2008) has argued many weaknesses of variable measurement in economics based on moments of probability, mean, standard deviation, skewness, and kurtosis. The author stated that most conventional methods fail to capture the “fat tailed” distribution of unlikely events, i.e., being correct 99%, which may be insufficient when consequences were large. And he observed that rare events will be infrequent or will not occur in most small samples (see Bye et al., 2011, for climate cycle composed of 30-year random walks, also Silver, 2012, pp. 190-193] for rare events). Therefore, this weakness also applies to the SPC chart, whose limits and moment calculations are based on small samples determined by the analyst. And even though the types of distributions are not limited, events that violate the rules, especially rule one, could include observations, which are many multiple standard deviations from the mean. As an example, Figure 1 below shows two different distributions plotted on the chart. The top image, the SP500 graph, shows limits of approximately +/- 2% on its Y-axis, while the limits of the bottom graph are approximately +30% to -35%. The red numbers visible near the observations denote each non-random occurrence, and the number itself represents the rule which was violated.

Mandelbrot and Taleb (2010) described these types of rare events in the bottom graph of Figure 1 as wild randomness, which they defined as any single observation or event that inordinately affected to the total. The authors gave examples of this type of concentration in certain winner-take-all markets. As examples, they cited best-selling authors by book volume or sales, internet traffic, and an outlier effect in the stock market: "10 trading days represented 63% of the returns for the past 50 years" (p. 50). In a table, the authors compared non-scalable versus scalable distributions. In the former, winners get small pieces of the total while in the latter, winners-take-all or the vast majority of the total. Non-scalable distributions include human physical qualities such as height and weight, where deviations are easily estimated. Scalable distributions are difficult to predict even from historical information and are governed by the "tyranny of the accidental" (p. 54). De Vany (2004) studied this phenomenon in the film industry and showed how only a few movies garnered the majority of revenue or were profitable while most generated losses.

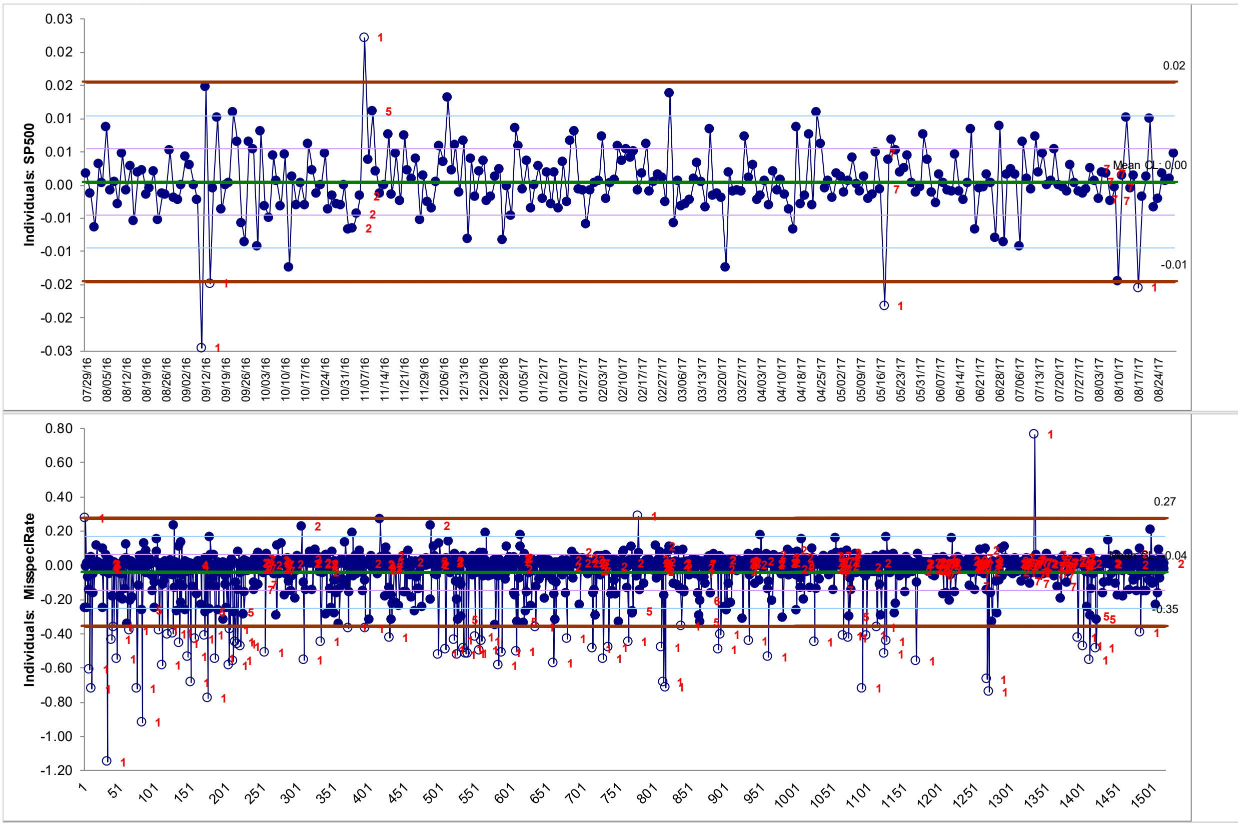

Figure 1. SPC returns for SP500 and a Sales Process Noting Scale Differences. Observations of two separate processes where the SP500 above indicates a plus or minus 0.02 upper and lower control limits with extreme values to -0.03. The sales process limits are 0.27 to -0.40, respectively, with extreme values to -1.20.

Characteristics Related to Type I and Type II Errors

Payoff curves. Taleb (2014) described two types of payoff curves: those that are concave with known visible gains and unknown errors, and those that are convex with small known errors and large possible gains. The latter were considered antifragile and based on optionality of future possibilities while the former was based on a narrative or story which might or might not fit the facts.

Similarly, in mapping types of fragility as a time series, Taleb (2014) distinguished between a negative blowup or black swan event after a series of positive gains and losses, which resulted in a catastrophic loss. This was described as fragility. Conversely, the positive value change resulting from a positive black swan would be antifragility. In between these two scenarios would be a robust investment which would be immune to both events. Gains with possible losses might be characterized as stock market investing, while losses to gains might characterized entrepreneurial start-up companies, venture capital endeavors, or sophisticated capital budgeting projects.

First mover advantage. Golder and Tellis (1993) examined first mover or pioneer advantage (FMA) across 50 categories of products and assessed the reliability of PIMS and ASSESSOR databases. Theories that favored FMA focused on the advantages of consumer loyalty and earning economic rents. However, the authors also offered several reasons why FMA could be disadvantageous such as free-riding the technology, changing consumer preferences, "incumbent inertia," or obtaining an ideal market placement (pp. 161-162). In reassessing FMA and the databases, the authors determined that market share leadership existed for 11% of 36 categories and that leadership was maintained an average of 12 years or a median of five years. Similarly, long-lived leadership occurred in only four of 50 categories.

VanderWerf and Mahon (1997) performed a meta-analysis of FMA. The authors found that using market share as the variable of FMA produced more significant effects than using profitability or survivorship. They noted, however, that market leadership was industry-specific and that there were cases of negative FMA.

Look what I did - look what I avoided. Taleb (2014) defined the idea of naïve interventionism, which was derived from a misunderstanding of the possible harmful effects of action. He observed that society recognized accomplishments appreciably more than avoidance of harm and that rewards generally came from things accomplished. Yet, he also described several examples of the benefits of procrastination: the Roman general Fabius Maximus who avoided direct fighting with Hannibal; the Fabian Society in the U.K.; the Latin expression *festina lente*, i.e. make haste slowly; or Lao Tzu's *wu-wei*, i.e., passive achievement. And he proposed that this bias for action was exemplified by the continued revisal of societal do's and don'ts based on the most recent statistical associations.

Similarly, Deming (1982) designed a Monte Carlo experiment to show how intervention behavior could be harmful. In his experiment, a funnel was placed at a given height above a target through which a marble is dropped onto a board. The marble would roll some distance away from the original target position. In the simulation, several interventionist rules are discussed which purport to maximize results. Rule one leaves the funnel fixed over the target position which is the optimal placement. Rules two through four create a compensating behavior (intervention) based on the previous marble end-position. For example, rule two moves the funnel target in the opposite direction of the final position of the previous; rule three moves the distance of rule two except it adjusts from the original target; and finally, rule four places the target over the last final position. The simulation demonstrates that rules two through four always increase the variation, and rule four transforms the series into a random walk. These tampering effects, at minimum, increase variation (rules two and three) or, at maximum, blow up a process (rule four). An actual example of rule two would be adjusting a quota to reflect current output; of rule three would be to increase bets to cover losses; and of rule four would be to allow workers to train their replacements in succession (SPC Excel [n.d.] for these and other examples;

Deming, 1982, pp. 331-332; Walton, 1990, p. 163). Similarly, rules two through four are designated as over-controlling a process, which is a typical type I error (SPC Excel, n.d.).

Personal fear of invalidity. Thompson, Naccarato, Parker, and Moskowitz (2001) described the history and various applications of the personal need for structure (PNS) and the personal fear of invalidity (PFI) constructs. In a brief history of the constructs, the authors stated that the desire to create meaning by understanding his/her surroundings was common among humans. They documented that Pierce (1877/1957) had proposed that this need was formed from the desire for survival and from the necessity to separate pleasant and unpleasant experiences. The authors also remarked that Heider observed that human desire was centered on achieving a particular state and on reducing doubt and uncertainty.

Since these constructs were developed from a basic human desire for structure and to deal with uncertainty, they are similar to each other and yet different constructs (Thompson et al., 2001, p. 22). Past studies had reviewed both constructs together or with others (Thompson et al. [2001] for PNS, PFI scale development; Clow & Esses [2005] for PNS, PFI, and need for closure [NFCC]; Rietzschel, De Dreu, & Nijstad [2007] for PNS and PFI; and Blais, Thompson & Baranski [2005] for PNS, PFI, and Need for Cognition [NFC]). But more often, many researchers in past studies have preferred the use of PNS as an individual variable (Svecova & Pavlovicova, 2016; or Davidson & Laroche, 2016).

In developing the two scales, Thompson et al. (2001) described PNS attributes as those individuals who prefer structure and clarity while avoiding ambiguousness and uncertainty. Also, these individuals would be seen by others as decisive and confident, and they would be perturbed by persons who were indecisive or vacillated with decisions or events. In particular, the authors noted that this behavior was generally praised in society but could lead to rigidity and a reliance on stereotypes.

Similarly, Thompson et al. (2001) noted that instead of concern for structure, individuals might be more concerned with the damage of committing errors, a high PFI. These persons were usually highly concerned with the risk of an undertaking, vacillating between courses of action, seeking alternatives, and exhibiting some distress when personal errors were enumerated. Moreover, the authors found a significant relationship between PFI and authoritarianism but an insignificant relationship to the rigidity of personal habits. Also, they found a strong correlation to depression, both public and private self-consciousness, and social anxiety.

Rietzschel, De Dreu, and Nijstad (2007) discussed previous studies of PNS and creativity, where higher levels of PNS generally had a negative effect on creativity. The authors studied the interaction between PNS and PFI and proposed that individuals who were high on both scales were likely to a negative relationship to creativity; however, higher levels of PNS with lower levels of PFI were expected to produce structured responses in uncertain environments. Moreover, they indicated that PNS resulted in simplification while higher PFI resulted in complexification. And the authors concluded that high PNS alone might allow for creativity, but coupled with high PFI, it might reduce it. Also, the authors did not find any significant relationship between PNS, PFI, or their interaction in the measure of flexibility (distinct semantic categories).

Similarly, Clow and Esses (2005) listed several significant findings for PFI while examining that construct with PNS and NFC. The authors found that when individuals created group stereotypes from descriptions, those with high PFI sought additional information but were less confident in their descriptions. When presented with a fixed number of descriptions to construct stereotypes, lower PFI individuals also created more detailed stereotypes. And when subjects were presented with a quantity of descriptions that were more or less than desired, high PFI and lower PNS and NFC individuals created more accurate stereotypes.

To review PNS, PFI, and NFC, Blais et al. (2005) used three comparative judgment tasks, which included assessing knowledge, vocabulary, and perception. While subjects with higher NFC had a superior performance on judgmental tasks, substructures for PNS and PFI did not show any significant effects.

The desirability of control. Burger and Cooper (1979) developed the construct of the desirability of control (DC). The authors conceived of the construct based on the work of Adler (1930), Kelly (1965), and Kelley (1971). A person with high DC wants to have control in their life. They prefer to influence others, often seek leadership within groups, and appear decisive and assertive. Where payoffs for control were small, both high and low DC individuals are not expected to exert control. Within a betting game, the authors found that high DC individuals exhibited an IoCON with pre-bets using a gambling scenario. The authors also reported several studies that investigated learned helplessness, which is manifested in individuals who are continually exposed to negative and aversive events. Other research suggested that higher DC individuals were more susceptible to learned helplessness.

Dembroski, MacDougall, and Musante (1984) tested the relationships between DC, locus of control, and other standard psychological measures, including the type A coronary prone behavior pattern. The authors made their type A assessments through both an interview process and quantitative measures. They found that high DC individuals were susceptible to the type A coronary prone behavior pattern, whether assessed through an interview or self-report. Furthermore, in the qualitative assessment, the authors reported significant relationships between interview-based type A subjects, DC, and voice mannerisms. However, no subjects who used the LOC I/E questionnaire showed any significant correlation with either the interview type A assessment, with voice, or with DC.

Gebhardt and Brosschot (2002) tested three separate samples of more than 300 respondents to assess the DC construct, to create a Dutch version, and to review the relationships with DC and LOC, coping style, repression, achievement motivation, personality characteristics, trait anxiety and depression, trait worry, burnout, and somatic complaints. Each construct was measured with its particular scale. Using factor analysis, the authors determined three subscales within DC, which included controlling others, relinquishing control, and self-control. While the authors listed 17 significant relationships to DC, the strongest positive associations were dominance, self-esteem, and problem solving, while the strongest negative associations were social inadequacy and avoidance strategies. Similarly, the authors also found significant negative relationships between DC and anxiety, depression, and worry.

True-False guessing. Developing a true-false guessing strategy was first documented by Fritz (1921). As an instructor, he noticed a pattern with students' incorrect responses as his pupils seemed to guess a true response more than a false one. To test his idea, he examined 19 tests among four professors and found similar results. To further investigate the phenomenon, he designed a 58-statement examination based on items found in a medical encyclopedia and devised questions which would be unknown to a lay-person. In two trials, the author obtained similar findings, i.e., 60.9% and 62% of the responses were true. Similarly, Poundstone (2014) described various true-false test-taking strategies and revealed that "recalling a fact was easier than creating a falsehood." Therefore, based on his research, he suggested that examiners were more likely to create true questions with a ratio of about 56% true questions to 44% false.

Krueger (1933) used a frequentist approach to assess several correct answers in a true-false guessing strategy based on question series of 10, 20, 30, 40, 50, 60, 80, 100, 200, 300, and 500 items using 1,000 subjects for each series. Within each of the sequences, a 50-50 representation of true-false questions was constructed. He also separated the subjects into two groups: one group who would know the 50-50 ratio and the other who would not be given any information. In the results for the 10-item questionnaire, seven of the subjects were able to guess either 10 or 0 correctly, while 31 subjects were able to get 9 or 1 correct. In the 20-item questionnaire results, no subjects were able to guess 19 or 20 questions correctly or 0 to 2 correctly; however, 4 subjects guessed 17-18 correctly. Therefore, a standard 20-question guessing questionnaire would severely eliminate the possibility of correct/incorrect guessing given the extreme values whose probabilities of occurrence are $(0.5)^{20}$ or 9.54×10^{-7}.

Similar to Poundstone (2014), Burton (2005) described 15 myths related to true-false and multiple-choice questions based on un-speeded tests, which allow respondents ample time for completion. While discussing those myths, the author elaborated on blind guessing, which he suggested was not totally blind. He observed that responses were not necessarily "completely uninformed and random" if clues were available from poor question construction or linguistical choices. Furthermore, the author commented on the difference between incorrect knowledge and complete ignorance and whether or not any differentiation should exist when scoring responses. He did not equivocate and emphasized that there was no difference between accepting or trusting misinformation and complete ignorance as any error based on false confidence (i.e., type I) might contradict the correct understanding of an entire topic.

Using a mathematical approach, Morrison (1978) suggested two possible "true" outcomes of true-false guessing. Using the example of a hypothetical taste test, e.g., Coke/Pepsi, or expensive/cheap wine, he observed that a subject either could discriminate between two options or would be forced to guess. The author compared the actual distributions to two other extreme outcomes: samples of 100% guessers and 100% discriminators. Through an iterative process and various statistical techniques, the author showed how the separation of guessers and discriminators could be determined.

And finally, distinguishing true-false has entered the medical field through experiments examining the brain using MRI techniques. Cabez et al. (2001) examined brain images of subjects while employing recognition tests of true-false. Their findings suggested that subjects were generally capable not only of recognizing true items and eliminating new items but also of accepting false items as true. The results also yielded true-false distinctions in the posterior medial temporal lobe (MTL).

Cohen's Kappa. Viera and Garrett (2005) provided a synopsis of Cohen's Kappa, which attempts to standardize the subjective judgments between observers of some phenomenon. The statistic is calculated by counting the agreement among the observers and agreement by chance, i.e., observed and expected agreements. There are four zones that require computation. Zones A and D count were both observers either agree positively or negatively. Zones B and C count, where either observer disagrees with the other. In a table the authors provided an interpretation of the statistic which ranges from -1 to +1: Kappa statistics less than 0, less than chance agreement; between 0.01 to 0.20, slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; 0.81 to 0.99, almost perfect agreement; and 1.0, perfect agreement. They also suggested that the statistic might be inappropriate for rare observations, which yielded low kappa values with high levels of agreement (also McHugh, 2012, p. 279). Feinstein and Cicchetti (1990) outlined the paradox of Cohen's Kappa by exploring the symmetry and balance between agreements and disagreements and demonstrating where those marginal totals might yield unexpected Kappa values. In a separate article, Cicchetti and Feinstein (1990) reviewed other omnibus indices to correct for these asymmetries and suggested that kappa should also be accompanied by individual values of agreement/disagreement. And McHugh (2012) argued for using both kappa and only the percentage of the agreement to enhance interpretational clarity.

Byrt, Bishop, and Carlin (1993) suggested an additional correction to the Kappa statistics by adding an index for bias and prevalence so that additional information could be gained. The authors recommended that the bias-adjusted kappa (BAK) should replace actual counts in zones C and D with average values. They also introduced two additional adjustments. The first was the prevalence index (PI), which was computed as the difference between the probabilities of yes and no. And the second was the prevalence and bias-adjusted kappa (PABAK), which would allow for a linear relationship in observed agreement.

Other research efforts have suggested further adjustments to Cohen's Kappa. For example, Kvålseth (2015) suggested a logistic transformation of negative Kappa values. Warrens (2014) provided new category coefficients for use with more than one rater. And von Eye and von Eye (2005) reviewed disagreement zones, triangles of disagreement, and disagreement by one unit on ordinal scales.

Summary

This literature review began with the history of randomness and examined how attribution of its effects was interpreted as the decision of the gods. Often these ideas were false. When discussing pattern recognition itself, various theories have suggested how the mind and memory function to make sense of our reality. In some instances where there are repeatable feedback loops, ideas of tacit knowledge and recognition primed decision-making show how experience can save life using historical behavior with the occurrence of new situations. However, psychology, in studying human subjects, has found extensive human bias or non-logical thinking with uncertain situations or in cases of limited feedback. For those of us working in a quantitative area such as finance and accounting, its mathematical nature tends to assign a scientific rigor that may not be deserved, and these facts were explicitly documented in the literature. Yet perhaps these weaknesses are not well-known to practitioners or those outside the profession. Several maladaptive behaviors showed how professionals could be deceived within their minds. So, perhaps many readers were surprised or appalled at these biases where others were dismissive of their importance. The latter may be those individuals subject to these illusions of the mind.

Furthermore, seeing these follies in other domains should reinforce our desire to accept the limits of our professional knowledge, which should not be judged as free from error, as is often the case. Concurrently, embracing those errors through self-critique can also advance our profession. As has been duly documented, the stakes are high, whether the outcomes are either positive or negative. Ironically, the inability to recognize, accept, and to profess these limits render many financial techniques no better than our ancestors' astragali, *I Ching*, or magic eight ball which is exactly where this literature review began.

CHAPTER 3 – METHODOLOGY

Proposed Model and Relationships

This study attempts to assess participants' type I and type II error tendencies using random sequence perception, as suggested by Fenton-O'Creevy et al. (2003). Furthermore, as proposed by Nickerson (2002), the research allows individuals to show their comprehension or instincts about randomness in a candid way by permitting them to write about their strategies, thought patterns, or ideas directly to the researcher through open-ended questions.

The proposed model for this study is shown in Figure 2, which delineates the types of positive and negative relationships between latent constructs and hypotheses. Each of the blue nodes or circles represents a latent construct. All extended literature elements and details of each latent construct and measure are found in chapter 2. The research expectations include desirability of control (DC) which is positively related to the type I error rate, and personal fear of invalidity (PFI) which is positively related to the type II error rate. Similarly, the characteristics are designed to show a positive effect on one type of error. In contrast, the polar opposite characteristics would indicate a positive effect on the converse error type. Each of these effects is expected and explained with hypotheses in the next section.

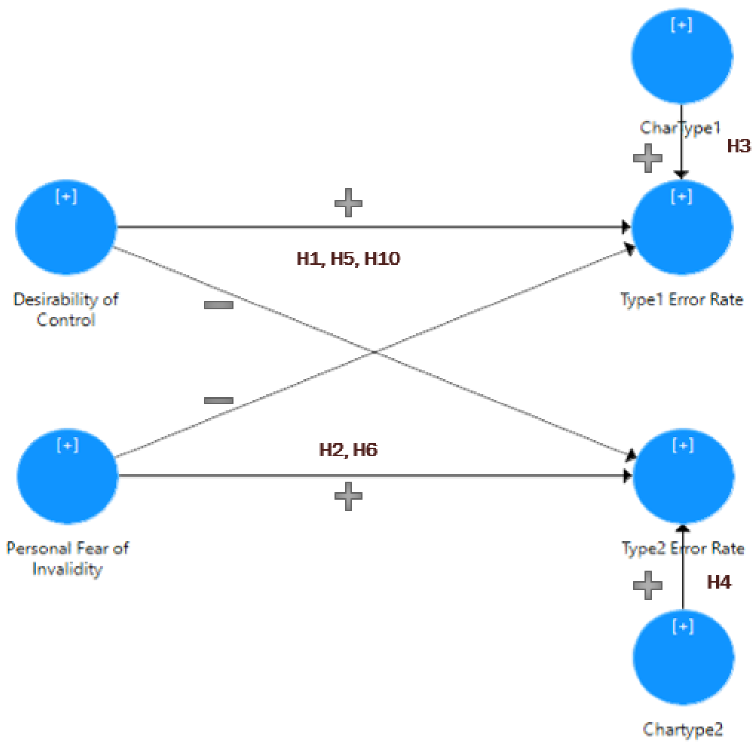

Figure 2. Proposed Model of Study and Associated Relationships. Latent constructs are blue circles with arrows indicating the direction of relationships. Hypotheses are numbered on the path of the appropriate relationship.

Hypotheses

Figure 2 shows how each of the ten hypotheses is related to the latent constructs within the PLS-SEM model. Hypotheses H1 through H6 and H10 are related to the PLS multivariate model, while hypotheses H7 through H9 are not SEM model-related assessments but are evaluations of the attributes of the sample. These particular hypotheses are related to the nature of the data obtained and whether the current sample maintains specific aspects from the results of historical studies.

Again, as seen in Figure 2, the desirability of control (DC) was anticipated to have a significant positive effect on the type I error rate. Consequently, the desirability of control, following Burger and Cooper (1979), would lead to an illusion of control where individuals would make type I errors. Therefore, *Hypothesis 1: High DC provides a significant positive path to type I error rate.*

Conversely, personal fear of invalidity (PFI) would have a significant positive effect on type II error rate. Individuals with type II error tendencies would want to avoid making any errors and might prefer the complexification of difficult questions. Consequently, those with high personal fear of invalidity, following Thompson et al. (2001), would defer to making specific pronouncements, add qualifiers, subtleties, and tend towards type II errors. Therefore,

Hypothesis 2: High PFI provides a significant positive path to type II error rate.

Certain behavioral characteristics are associated with both type I and type II errors. These sets of characteristics are mirror images of each other and are expected to have significant positive effects on each error type. The first characteristic, payoff curves, following Taleb (2014), suggests that gains with possible losses characterize typical stock market investments while robust and anti-fragile investments might offer small losses with the possibility of large gains. The second, first-mover advantage, following Golder and Tellis (1993), argues that first movers may not actually be first in the market or are perhaps are those that were first remembered, i.e., survivorship bias. Third, a more subtle characteristic examines an avoidance tendency versus an action orientation in individual behavior. Any action certainly opens the opportunity for type I errors, following Taleb's (2014) discussion of a bias for action. And as previously explicated, Deming (1982) warned that action on any process would be considered tampering and would worsen outcomes. However, individual incentives and compensation within an organization do not usually reward people for avoiding errors. Therefore,

Hypothesis 3: Characteristics of type I (gains to losses, first-mover advantage, look at what I did) provide a significant positive path to type I error rate.

Hypothesis 4: Characteristics of type II (losses to gains, follower strategy, look what I avoided) provide a significant positive path to type II error rate.

Furthermore, an assessment of true-false guessing strategies would be explored. Cohen's Kappa, an assessment of interrater reliability, is used to form groups of respondents from that assessment following Viera and Garrett (2005). Based on simulations with extreme oscillation bias, at least two select groups should emerge either by chance, tacit tendency, or specific individual strategy in sequences of true-false or false-true, following Poundstone (2014): those with a propensity to commit type I errors and those with a propensity to commit type II errors which are positively related to each respective error type as in Figure 2. Therefore,

Hypothesis 5: Kappa group of type I from the True-False assessment provides a significant positive path to type I error rate.

Hypothesis 6: Kappa group of type II from the True-False assessment provides a significant positive path to type II error rate.

Within the true-false assessment responses as mentioned above, some participants were expected to make excessive non-random alternation (under-alternation or oscillation) between true and false answers following Ayton and Fischer (2004), Poundstone (2014), Lopes and Oden (1987), and Bar-Hillel and Wagenaar (1991). Therefore,

Hypothesis 7: Because of heuristic bias, the alternation rate of actual True-False assessment will be significantly different from a randomly generated True-False assessment.

Both the endogenous error rate responses and the exogenous true-false assessment responses are evaluated by calculating the percentage of correct responses and rates of type I and type II errors. Logically, if respondents could distinguish between an actual deterministic series and a random one, the percentage correct in the SPC error rate evaluation should be superior to respondents' guessing strategy. Therefore,

Hypothesis 8: Percentage correct of the SPC Error Rate assessment is significantly higher than the percentage correct in True-False assessment.

Following Fritz (1921), Krueger (1933), Burton (2005), and Poundstone (2014), respondents would answer a higher percentage of true responses to false responses. And when answering true, the probability of making a type I error would also increase. Therefore,

Hypothesis 9: Distribution of the actual percentage of True responses in the True-False assessment group will be significantly higher than the percentage from a random distribution.

Hypothesis 10: a higher percentage of True responses in the True-False assessment group provides a significant path to type I error rates.

And lastly, the qualitative assessment explores the strategies employed, if any, in the assessments of both the error rates and true-false latent constructs.

Justification of Methodology

Following a pragmatic research view, this study used a concurrent mixed-method approach, i.e., data were collected at one time and contained both qualitative and quantitative elements. The mixed-method approach is appropriate as it attempts to uncover new research directions, as suggested by Fenton-O'Creevy et al. (2003) and by Nickerson (2002) as previously explained. And as advocated by Hussein (2009), using mixed-methods can enhance an area of study by triangulation of the underlying phenomenon.

For the quantitative portion, a multivariate design was employed, which includes the use of structural equation modeling using a partial least squares method (PLS-SEM). Buhl, Goodson, and Neilands (2007) recommended the use of SEM based on four factors. The first, based on the complexity of many research questions, favored SEM as it can handle many variables. Similarly, SEM facilitated the examination of multiple relationships and accounted for experimental error better than older multivariate methods. The method also offers data analysis flexibility because it includes the calculation of all effects (total, indirect, and direct) and is more forgiving with the treatment of incomplete data versus the older techniques. Lastly, Astrachan, Patel, and Wanzenried (2014) stated that more indicator variables were retained with PLS-SEM than with CB-SEM.

For the qualitative portion, Creswell (2009) advised several types of mixed method strategies available to researchers. This study uses the embedded design where the qualitative data is collected simultaneously and plays a supportive role to deepen and enrich the analysis. The open-ended questions would allow individuals to show their comprehension or instincts about randomness candidly by permitting them to write about their strategies, thought patterns, and ideas directly to the researcher.

Population and screening criteria. Within the behavioral finance literature, rarely have actual practitioners been surveyed. For example, Fenton-O'Creevy et al. (2003) created experiments to assess actual traders' susceptibility to IoCON and their performance in market outcomes within an organization. Similarly, real accounting professionals were used by Uecker and Kinney (1977) to test for sampling errors; by Biggs and Wild (1985) to review estimates of unaudited results and investigate bias, and by Johnson (1994) to test the memory of audit evidence between individuals and groups. Taleb (2004, 2007, 2014) discussed bias in financial and economic industry behavior generally while Heuer, Merkle, and Weber (2016) studied actual investor misattributed behaviors. However, most psychological studies use student populations as participants and subjects, and they may lack practical experience in the workplace.

For this study, corporate finance and accounting personnel are surveyed. Screening criteria employed were based on business functional areas such as marketing/promotion, customer service, sales, accounting/finance, distribution, R&D, administrative/management, production, operations, IT, purchasing, and legal. From these groups, only finance/accounting are selected to continue to the survey. If an individual had worked in different functional areas, respondents were requested to choose their functional area based on the highest number of years toiled. The level of accounting/finance experience was a minimum of five years following Johnson (1994). A bipolar 6-point Likert scale was used to evaluate a tendency towards qualitative/quantitative skills. Furthermore, consideration was given to stratified samples based on qualification level, job title, company size, and/or international experience. Due to budget considerations, none of these criteria were used for screening; however, the first two items were collected for informational purposes only.

Data collection and sample size. An initial survey instrument was created, which included 84 observed variables, 11 latent constructs, and the screening variables. Kyriazos (2018) discussed the several factors affecting SEM sample size, which included model complexity, type of data distribution, reliability of the underlying variables, missing or imprecise data, number of interactions, among other criteria. The following parameter estimates were used to determine minimum sample size for this study: minimum effect size of 0.20 based on previously unpublished research (using a 0.30 would not change in minimum respondent count); a standard power level of 0.80; the latent and variable counts previously mentioned; and a standard probability level of 0.05. These combined factors yielded a minimum sample size of 588 based on the model structure and 488 to detect an effect.

Participants for this study were reached through a screened Qualtrics panel of respondents. During a previous research project, the Qualtrics database of potential respondents allowed for an ample dataset that was returned to researchers punctually. Using these participants meant that this study was sourced from a voluntary response sample. The estimated time to completion of the survey instrument was about 20 to 25 minutes.

Measures

Each of the measures in this study was separated into the endogenous, exogenous, grouping, and demographic variable categories. The exogenous variables are those exterior to the model and similar to independent variables, while endogenous variables are those being studied and are internal to the model.

Endogenous variables. The endogenous variables, rates of type I and type II errors, were created from the results of a 20-question sequence instrument constructed using statistical process control charts to assess RSP and error rates. The choice of 20 questions follows Krueger (1933). A data set, crafted using a random number generator with a specific mean and standard deviation, is used to create the charts. From this sequence, ten strings were selected that were random, and ten strings were selected that violated the selected random rules used in SPC charts, as shown below in Figure 3. Twenty separate sequences-questions were generated from the dataset with each non-random error being included at least once to create a 50-50 ratio of random to deterministic items (see Table 5 for a list of non-random rules). Therefore, this assessment was a formative latent construct consisting of true-false binomial responses (random/non-random) that generated a percentage of correct answers, type I and type II errors. Calculating these rates transformed the constructs into continuous measures.

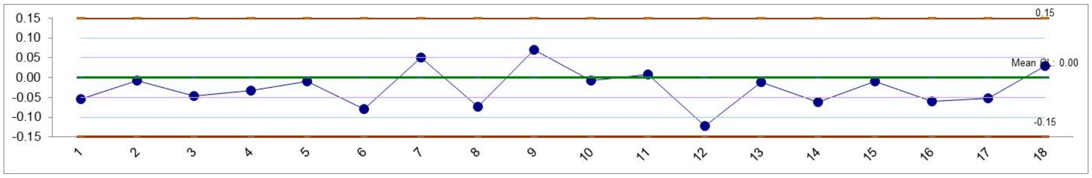

Figure 3. SPC Chart Randomness Diagram
Eighteen observations from the average or green centerline. Brown lines show control limits of three standard deviations.

Consideration was also given to include the actual data points of each event or use only a graphical representation. For learning and education, Schnotz and Wagner (2017) suggested that both pictures and graphs were superior to either case. However, when examining economics statistical and error terms, Soyer and Hogarth (2012) found that economists provided more accurate interpretations with graphical representations only. To avoid unnecessary complexification, which might also generate distraction and given the selection of a speeded test to be discussed below, graphical representations were considered the better choice.

Initially, two scenarios were considered to frame the test using one accounting and one finance context. But, in the interest of simplification, one scenario was selected using financial references based on abnormal stock returns. The exact wording of the scenario is as follows:

> For a company automation project, you are a trader reviewing a fictional stock for daily abnormal returns after a series of 18 trading days. Abnormal returns occur when the series of percentage changes from one day to the next is not random. A random series of returns do not exhibit any patterns. You must identify those sequences or series of returns as either random or non-random. You will be asked to view 20 different individual scenarios.

Specific details on the scales, standard deviation, and limits calculations for the SPC chart, as illustrated in Figure 3, were explained in detailed instructions to respondents without hinting about the nature of any specific random rules following Burton (2005) and Ayton, Hunt, and Wright (1989).

Consideration was given as to whether answering time limits, which created speeded tests, would be preferable to un-speeded ones. In this approach, a speeded test was preferred because of two factors: first, the assessment forces the respondents to answer correctly or incorrectly without being pre-informed of the questions, and second, the timed element eliminates the possibility for internet or other types of searches. For these questions, 25 seconds per question was initially adopted.

The validity and reliability of the SPC chart are thoroughly discussed in Chapter 2. It was demonstrated how this tool separates common variation from special cause (random versus non-random) along with the probability of each of the eight rule occurrences occurring randomly. The history and continued use of this tool by many businesses and the unceasing development of its software point to the robustness of this application.

Exogenous variables. The two exogenous latent constructs use standard psychological scales. The first, desirability of control (DC), follows the development by Burger and Cooper (1979). The scale is composed of 20 questions on a 6-point Likert scale of "this statement doesn't apply to me" to "this statement always applies to me." Two sample statements include "I prefer a job where I have a lot of control over what I do and when I do it" and "I try to avoid situations where someone else tells me what to do." And of the 20 items, seven are reverse coded. For this study, DC is a reflective, ordinal latent variable.

Burger and Cooper (1979) performed as an assessment of DC reliability using the initial sample of respondents and obtained a Kuder-Richardson value of 0.80. And, with a second sample, the authors obtained a coefficient of 0.81. Similarly, in creating a Dutch version of the DC scale, Gebhardt and Brosschot (2002) received an overall Cronbach alpha of 0.77. Furthermore, Burger and Cooper (1979) got a test-retest coefficient of 0.75 after a six-week interlude between assessments. And after a one-week interval, Gebhardt and Brosschot (2002) obtained a test-retest coefficient of 0.86.

Burger and Cooper (1979) also demonstrated the discriminant validity of the DC scale. The authors simultaneously administered the locus of control test to over 200 in the original sample. A small negative relationship between DC and LOC was found of -0.19. Gebhardt and Brosschot (2002) obtained mixed results of discriminant validity as they tested both multi- and unidimensional versions of LOC. The unidimensional scales were positively correlated but not significant overall; however, the multi-dimensional version was significantly positively correlated with an internal LOC subscale and significantly negatively correlated with "powerful others" LOC and "chance" LOC subscales. And finally, when Dembroski, MacDougall, and Musante (1984) separately tested low DC and high DC groups with LOC, they found neither significant differences between groups nor any significant correlation between DC and LOC. The authors detected a small significant correlation between DC and two type A scales.

The second exogenous variable, personal fear of invalidity (PFI), follows Thompson et al. (2001). The scale is composed of 14 questions on a 7-point Likert scale of "strongly disagree" to "strongly agree." Two sample statements include "I may struggle with a few decisions but not very often" and "I never put off making important decisions." And of the 14 items, five are reverse coded. In this study, PFI is a reflective, ordinal latent variable, and the

Likert scale anchor points will be adjusted to a 6-point Likert to follow the DC scale (for researcher preference and reliability see Chyung, Roberts, Swanson, & Hankinson [2017]; also, Garland [1987]; Armstrong [1987]).

Thompson et al. (2001) assessed PFI reliability using the initial sample of respondents and obtained a final Cronbach alpha of 0.84. And with a second sample, the authors obtained a coefficient of 0.94. Similarly, Clow and Esses (2005) reported a Cronbach alpha of 0.82.

Thompson et al. (2001) also demonstrated the discriminant validity of the PFI scale. The authors simultaneously administered the right-wing authoritarianism (RWA), rigidity about personal habits (RAPH), and the Beck depression inventory (BD) scales to 157 participants in a previous round. Small positive significant relationships between PFI and RWA, and PFI and RAPH were found of $r = 0.22$ and $r = 0.15$, respectively, which the authors cited as evidence for discriminant validity. A large positive significant relationship between PFI and BD was found of $r = 0.64$, which the authors cited as evidence for convergent validity.

Four exogenous characteristics that suggest a behavioral separation between individuals with either type I or type II error tendencies were identified. This variable is composed of five questions on a bipolar 6-point Likert scale. Two statements concern the types of payoff curves, following Taleb (2014), and are anchored by "small losses with a small probability of a large gain" and "small gains with a small probability of a large loss" or by "medium losses with a medium probability of large gains" and "medium gains with a medium probability of large losses." One statement concerns first-mover advantage, following Golder and Tellis (1993), and is anchored by "follower strategy" and "first to market, first-mover strategy." The final two statements concern tampering and interventionism, following Taleb (2014) and Deming (1982), and were anchored by "look what I avoided" and "look what I did" or by "trial-and-error" and "tried-and-true." The left-side anchors reflect the tendency towards a type II error while the right-side anchors reflect type I leanings. Specifically, the type II scale reflects the mirror image of the type I.

Grouping variables. With this project, two types of grouping variables were selected for comparison within the SEM model. The first is based on a true-false guessing assessment in which questions were designed to be unanswerable, as in Fritz (1921) and Krueger (1933). A 20-question assessment was created based on a random selection of certain words within several personal book selections and was designed to be unknown to respondents. Again, the choice of 20 questions follows the error rate assessment. Two sample statements include "Qohelet 1:5, Chouraqui Bible; 8th word is 'son,'" and "Qohelet 1:5, Chouraqui Bible; 8th word is 'lieu.'" Each question follows this method where each pair of questions was constructed using precisely the same wording except for a "true" word or a "false" word highlighted in quotes. This assessment is a formative latent construct consisting of true-false binomial responses. These responses will generate percentages of correct answers, type I, and type II errors, which mimic the endogenous error rate responses.

The true-false grouping variable was based on Cohen's Kappa calculation between respondents of the true-false assessment. As mentioned in the hypothesis section, simulated responses with excessive oscillation bias would create at least two self-forming groups. It was expected that no more than 40% of the respondents would have unique answers. Similarly, the second grouping variable was based on the number of true responses as respondents were expected to answer about 60% true following Poundstone (2014), Burton (2005), Krueger (1933), and Fritz (1921).

Similar to the SPC chart question construction, both speeded, and non-speeded tests were considered. For the same reasons as previously discussed, a speeded test was preferred. To maintain this section of the survey to force guessing only, a 15-second limit was selected.

Two statements were also created to assess firm or familial stress as major past or current events, which could have affected individual responses toward error types, following Kriz (1993). These questions employed an "affected me" 6-point Likert scale anchored by "not at all" and "significantly." These two statements were as follows: "How much has participating in a financially distressed company affected your business perspective? (Financial distress includes major reorganization, credit workout, rescue, bankruptcy, and liquidation)" and "How much have particular life stresses affected your business perspective?"

Demographic variables. Several demographic variables were selected for evaluation. Generational information was gathered based on birth year ranges, as described by Dimock (2019). Zysberg and Kimhi (2011) had noted some differences within gender, as did Burger and Cooper (1979); therefore, male/female gender was collected. Even though Fenton-O'Creevy et al. (2003) had not found any differences between education and IoCON, the level of education was also added to the survey. These demographics questions use a five-category ordinal response scale for birth year, from the Millennial to the Greatest generations; a three-categorical response for gender; and a six-categorical nominal response for education, from some high school to a graduate degree.

Qualitative variables. Two open-ended questions were created to understand various strategies that respondents might employ to answer both the true-false guessing and the error rate assessments in the event they had no knowledge. The suggestion to directly ask respondents follows Nickerson (2002), while Morrison (1978) provided a method of iterative estimations to separate guessing behavior from actual knowledge. The two sample statements include, "Please discuss how you determined your answers to the graphical series charts (statistical process control) test questions. All details about your thought processes are appreciated." and "Please discuss how you determined your answers to the true-false test questions. All details about your thought processes are appreciated."

Analysis Plan

Following Christensen, Johnson, and Turner (2014), several pilot surveys for the questionnaire were undertaken to receive feedback on the instrument. Several areas were verified to ensure clarity of understanding among respondents especially concerning the financial scenario, the SPC graphs and their visibility and size of the display to the participant, the open-ended questions designed to elicit participant strategies, the timing length of both the true-false and SPC un-speeded assessments, and the overall instruction guidelines. A convenience sample of colleagues in the accounting and corporate finance area, and friends were asked to participate without any of the screening data. Feedback was used to modify it as necessary.

Hair, Hult, Ringle, and Sarstedt (2017) outlined an eight-stage procedure for the application of partial least squares structural equation modeling (PLS-SEM). The first step involves specifying the structural model, creating the latent constructs, and their respective relationships. The structural model or inner model shows the directional relationships among latent constructs (blue circles), as displayed in Figure 2. In step two, the measurement or outer model is specified, which denotes the relationships between the measures and each latent construct. In Figure 4, DC is a reflective variable, and the characteristics of type I errors is a formative variable. The arrows of the reflective variable are directionally away from the latent construct and towards the individual measures while the arrows of the formative constructs return towards the variable.

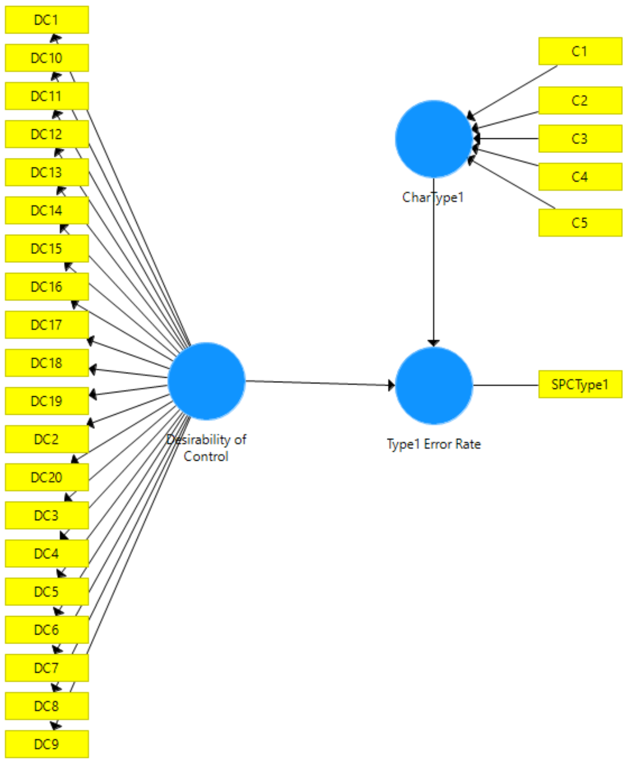

Figure 4. Example of Reflective and Formative Constructs.
A partial view of the internal model showing formative, reflective, and single relationship variables.

Step three involves data collection and the analysis of the initial responses for the treatment of missing data, suspicious response patterns, or outliers. In step four, the model calculation begins with the PLS algorithm, which assesses path model, i.e., beta regression weights between constructs, and both the outer weights for formative constructs and outer loadings for reflective ones. In step five, the outer measurement model is evaluated. For reflective measures, internal consistency, and convergent and discriminant validity were assessed by Cronbach's alpha and composite reliability; indicator reliability and average variance extracted; and Fornell-Larcker criterion, respectively.

Similarly, for the formative measures, convergent validity, the significance of outer weights, and outer loadings were assessed. For discriminant validity, the hetero trait-mono trait method (HTMT) is preferred, and for internal consistency, Cronbach's alpha and composite reliability are preferred. In this study, convergent validity was determined between the characteristics of error types and actual type I and type II error rates as no previous global measures exist. Collinearity was evaluated using the variance inflation factor (VIF), while a bootstrapping procedure would assess significance and relevance. Step six assessed the structural model itself by reviewing collinearity (the model VIF); the significance of the path coefficients through bootstrapping; the value of the coefficient of determination, R^2, and the effect size, f^2, which assesses the strength of R^2; the out of sample predictive relevance, Q^2, and the effect size q^2, which assesses the strength of Q^2. In step seven, moderation and mediation are analyzed but were not integrated into this model. And lastly, the final step involved all the interpretations and discussion.

Similarly, all the grouping and demographic variables are entered into the Smart PLS software as groups. Steps four through six were repeated for each of those variables to determine which groups are significantly using the same criteria as the complete model. The specific group model would be the same as in Figure 2, and interpretations were based on the identical complete model.

Moreover, several tests are designed to verify the historical relationships within the sample and to a randomly generated sample from a simulated true-false assessment. Three relationships were examined internally within the sample of respondents using simple t-tests: the percentage correct, type I, and type II rates. T-tests were used to evaluate the percentage of true within the true-false assessment and a randomly generated data set of binomial responses. And because of a bias towards alternation (excessive oscillation), a statistical runs test, the above/below non-parametric exact runs tests, was employed to assess excessive alternation or lack of it within both the SPC and the true-false assessments. This runs test reviews the number of singlets, doublets, triplets, etc. in a series for each type of binomial response, counts the actual responses, and creates a confidence interval to assess random versus non-random alternation. To avoid type I errors, a 99% confidence level was selected, which was recommended by SigmaXL statisticians.

For the open-ended questions, Creswell (2009) proposed several analytical steps that included data cleaning, developing codes and chunking, clustering them to find patterns, and summarizing the findings. *A priori* codes were deduced from the chapter 2 literature review of type I errors such as positive or negative recency, alternation, etc. but these potential strategies would need to be validated when reviewing respondent answers. Conversely, *a posteriori* codes could be induced from the actual data as well as respondents' psychological valence. Both of these methods were incorporated into this study. NVivo software was the means employed in the analysis because it offers many visualization tools such as word trees, frequency diagrams, network graphs, cluster diagrams, and word clouds, among others.

CHAPTER 4 – DATA ANALYSIS AND FINDINGS

This chapter encompasses the creation of the research instrument, the three pilot tests to assess various aspects of the questionnaire, the final view of the research instrument and its rollout to the Qualtrics panel of respondents, the qualitative and quantitative analyses, and the findings of this study in a discussion at the end of the chapter.

Survey Construction and Pilot Studies

After a successful proposal defense, the Institutional Review Board (IRB) of Rollins College approved this dissertation research after the submission of the required documents. In this section, questionnaire construction, pilot surveys, and feedback from Qualtrics are highlighted.

Questionnaire construction. Building the questionnaire was an iterative process that traversed many versions as new information was obtained from participants, which included the DBA cohort students, friends, and colleagues. The initial graphical representations of the 18-point SPC charts used to create the endogenous variables were produced from a random number generator using a mean of 0.0, a standard deviation of 0.05, and a sample of 100,000. These random strings created both random and non-random sequences by using the SPC chart assessment tool within the SigmaXL software. Ten representations of each were extracted from the sample and recalculated again using the software. Also, during the selection of the ten non-random strings, six of the eight Nelson's rules were selected one time while two rules were selected twice (rules one and two). Next, each of the questions was randomized and inserted into the instrument. No further randomization occurred internally within the survey because, in the case of guessing behavior, question order was important. This order would be lost if the questions were randomized for each recipient. Appendix Table A1 shows the question number, the answer, the error type if incorrect, and rule violation if applicable, in the final randomized order.

A similar process was undertaken for the creation of the true-false assessment. Various sentences were extracted from books in a personal library where the ten true questions were based on random word selection, and the ten false questions were created from selecting either the word before or after the correct one within the same sentence. Appendix Table A2 shows the question number, the answer, the error type if incorrect, and the actual question in the randomized order.

The exogenous variables, desirability of control (DC), and personal fear of invalidity (PFI) questions and their respective scales can be viewed in Appendix Tables A3 and A4, respectively. The question numbers, questions, and coding (straight or reversed) are listed for each. The characteristics' questions with bipolar scales are also listed in Appendix Table A5 and with the Likert scale items in Appendix Table A6. Lastly, Appendix Table A7 contains the demographic items selected for this study, which included generation, education, and gender.

All these items were incorporated into an initial Qualtrics questionnaire. During this process, it was discovered that "trial and error" versus "tried and true" were reversed on the bipolar scale. The former would be associated with type I errors (repetition) while the latter (non-repetition) would be correlated with type II. Specifically, "tried and true" implies that if maintenance, changes, or adoptions within a system never occur, then one expects that system to eventually decay and break-down at some future point. For "trial and error," overcontrol or the constant adjustment of any process, when unnecessary, would cause a type I error (see Deming, 1982).

Concurrently, as the questionnaire was assembled, additional phraseology of the qualitative questions was examined based on feedback from Rollins colleagues. The initial open-ended questions were changed from "Please discuss how you determined your answers to the graphical series charts (statistical process control) test questions? All details about your thought processes are appreciated" to "How do you think other individuals answered the graphical series test questions." The same verbiage was also applied to the true-false test questions. This style of wording was preferred based on the research of Grossman and Kross (2014), who discussed what they termed "Solomon's paradox," which stated that individuals make better decisions for others than they do for themselves. The authors reported that subjects showed improvement in their reasoning when thinking about others rather than themselves. So, when participants were asked to move away from an egocentric perspective to a distance away from themselves and towards another individual, the paradox was eliminated. Furthermore, the authors asserted that age played no role in superior reasoning, which is contrary to the wise-old sage archetype.

Initial pilot trial. Several pilot studies were completed to look for any issues related to the scenario, the instructions given to respondents and their interpretation of those instructions, the visibility and clarity of the SPC charts, the phraseology of the open-ended questions, and the timing length for both the SPC chart questions and the true-false grouping variable following Christiansen, Johnson, and Turner (2014).

The initial complete version of the questionnaire was delivered to a convenience sample of finance and accounting respondents from a major telecom firm. Two full questionnaires were completed, while one was only partially finished. In an initial analysis of the percentage correct for the SPC charts and the true-false, the pairs were 60%-50%, 50%-50%, and 35%-35%, respectively. To obtain a significant result from a 20-question test, 16 of 20 or 79% must be answered correctly (*p-value* = 0.049, effect size = 0.61); therefore, no initial results showed knowledge above guessing. Similarly, and as anticipated, two true-false strategies were observed: replying both all true or all false, which yielded a 50/50 correct-type I or correct- type II error, respectively. This would appear to be an optimal strategy to maximize the score of correct answers and to minimize a result which is less than random.

Average survey completion times ranged from 14 to 19 minutes. The average time to complete each SPC chart assessment was about 9 seconds with a standard deviation of 5.5 seconds. The coefficient of variation of 0.61 suggests a uniform distribution. The true-false questions averaged 2.5 to 4.5 seconds with a maximum of 15 seconds and a standard deviation of 2 seconds, which suggests, once again, a non-normal distribution.

The results of the qualitative questions were disappointing. Responses to "how do you think other individuals answered the graphical series test questions" included "similar to my answers," and "I think other people were just looking to see how drastic changes were to determine randomness." Moreover, responses to "how do you think other individuals answered the true-false test questions" included "most others may have created a mix of true and false guesses" and "randomly." Assessing the percentage correct, type I and type II rates did not exceed random results while testing for cluster, mixture, trend, and oscillation bias were not significant.

Each of the respondents was emailed to see if they would submit to a post-questionnaire interview which was based on Gigerenzer (2000). Within his "write aloud" protocol, he suggested three steps to understand whether or not a subject guesses: first, respondents would write an answer to a question; second, researchers would determine whether or not a guess or calculation was made based on that text; and third, an interview with each subject would occur to test the researchers' assessment methods. In this case, one respondent agreed to be interviewed while the second did not wish further contact.

The interview lasted about 30 minutes and covered all items in the questionnaire. Some of the topics included the following themes. The scenario instructions were deemed clear and comprehensible; however, he preferred "think I see" and "think I do not see randomness" versus a true-false response to the question, "Is this string random?" His interpretation of non-randomness within each string was to "look for trends." The initial time of 25 seconds for the SPC chart representations was perceived to be too short, which was confirmed in a review of the "time to page submit" calculation. His suggestion was 40 to 45 seconds. However, the true-false assessment time of 15 seconds was not considered limiting. The respondent confirmed that once he realized the nature of the unanswerable questions, he consistently responded true. Also, qualifiers for the bipolar scales of "small, medium, large losses or gains" were discussed. He suggested additional simplification.

Furthermore, the open-ended questions were again reviewed along with two new options based on Taylor and Small (2002) who, in a meta-analysis, had assessed differences in reliability between situational and past behavioral questions in structured interviews, i.e., employing a "suppose that" format versus a "can you think of a time when" format. Likewise, the interviewee discussed whether or not questions centered on distinguishing random events in life would be appropriate. Finally, some grammatical errors were also highlighted.

Second pilot trial. Because of the disappointing responses to the qualitative portion of the study in the first pilot and to provide additional insight into the respondents' perceptions of random sequences, a second partial trial to differentiate the types/styles of qualitative questions was warranted. Additional research was required to provide alternative open-ended approaches from which respondents could choose a preferred method. Allowing this choice would further enrich the quantitative portion of the study.

Beyond Solomon's paradox of Grossman and Kross (2014) and the write aloud method of Gigerenzer (2000), other qualitative methodologies were considered. Among them, Overgaard, Gallagher, and Ramsoy (2008) reviewed first-person data collection methods. The authors began with a history of these approaches but followed with descriptions of the two major trends in psychology today. The first, neurophenomenology, which is currently termed microphenomenology, is a method where subjects gain insights into their conscious experience through mental training. Furthermore, subjects are guided by an interviewer who, through open-ended questions and not predefined categories, aids them in recognizing their experiential knowledge and making it available to others. The second, front-loaded phenomenology, seeks to differentiate human actions between a sense of agency and ownership. The authors described the concept of agency as occurring when individuals were in control of a movement, i.e. its initiation and causation such as raising an arm; however, the concept of ownership occurred when, for example, a researcher would complete a movement for a subject, i.e. raise the subject's arm while the subject would remain passive even though the actual movement would indeed occur. Similarly, Kordes (2012) described a complementary approach called Descriptive Experience Sampling (DES), where the time between a stimulus and its description was minimized. For example, a subject carries a device on his/her person, and when an alarm sounds, the subject records his/her thoughts at that moment with as much detail as possible. Preselected categories are used to classify the descriptions which provide a better understanding of the lived experience (see also Petitmengin, 2006).

Despite the fact that microphenomenology is a qualitative interview method, it offers an avenue for open-ended questions. Petitmengin (2019) provided the main scientific criticism of the study of introspection in western cultures generally believing that it is impossible to actually do a task and simultaneously be an observer of that task. Is it possible to "walk in the street while watching oneself from the balcony," Auguste Comte summarized western philosophy (as cited in Petitmengin, 2019, p. 3)? However, the author noted that this method was successful in understanding implicit learning and tacit knowledge and had helped experts who did not understand certain hidden aspects of their knowledge. Once this knowledge was uncovered, it could be passed along to others, used for training, etc. For example, the firefighters studied in Klein's theory of recognition primed decision making (RPD) would be interviewed to delineate specific items of insight within specific experiences that led to a particular course of action. Similarly, Petitmengin (2006) described microphenomenology as a method which went beyond "what" in a narrative sense to the actual "how" in an actual sense. Similarly, she critiqued the think-aloud method of Ericsson and others as these types of methods would not ascertain the actual attribute that determined choice but rather default to a narrative description of what occurred (p. 241). Her description of the interview process occurred in five steps: stabilizing the subject's attention on an experience; focusing on the experience; reformulation when the subject digressed; returning to the actual particular experience; and encouraging the subject as he/she progressed in uncovering a conscious process (direct reference).

With these two additional techniques, a new pilot questionnaire was created to assess the three qualitative styles of questions. The survey began with the SPC scenario and requested comprehension and comments. Next, using three SPC charts, a question based on Solomon's paradox was constructed (Grossman & Kross, 2014): "How do you think other individuals answered the graphical series test question from the previous page?" Next, several questions based on microphenomenology were constructed (Petitmengin (2006, 2019): "Please think back to when you first viewed the graphical questions. What did you do?"; "How did you begin?"; "And then afterwards?" And for the third style, one question based on either write aloud (Gigerenzer, 2000) or think aloud (Ericsson, 2003) was created: "As you are viewing the graphical series question, please write down all thoughts, ideas, or internal mental commentary as they occur to you and which lead to your choice of random or non-random." A final survey question asked respondents to rank which type of question provided their best explanation. Then, this instrument was sent to another convenience sample of accounting and finance professionals.

Additionally, three true-false questions using the three styles mentioned above were sent to volunteers with the Rollins Gameboard Club. Since the initial pilot had revealed specific guessing strategies, this short survey sought to confirm preferred qualitative styles that would create additional depth of responses from individuals who specialized in gaming strategies.

For both evaluations, there were eight respondents, four from accounting and finance, and four from gamers. In the quantitative assessment, the percent correct for both groups was 33% and 50% for SPC and true-false, respectively. In the SPC group, one person scored all three charts correctly, while others correctly identified only one. In the true-false group, each person scored either zero to three responses correctly, which suggested large potential variation in responses. As noted in chapter 2, when individuals guess answers to questions, the expectation is roughly 60% true. Gamers guessed at a true rate of 67% or 2:1. However, within the SPC group, the result was 50% or 1:1. These outcomes were encouraging while realizing that the sample size was small.

In choosing a preference between the styles of advising others (AO), write-think aloud (WTA), and microphenomenology (MP), no respondent in any group preferred AO. In fact, this style produced three single-word answers, which would add little depth of understanding to the research questions. Within the SPC group, two preferred both WTA and MP, respectively, while in the true-false group, three out of four preferred WTA. In character (letter) counts within each text box, AO consistently and in all cases produced the shortest responses, but in five of the eight responses, MP produced longer ones. This respondent perhaps captured the reason for WTA preference: "The question number two option which was like why did you choose this, it was the most helpful because it was on the same page as the graph," i.e., being able to see and write concurrently.

Several other results were confirmed. The presence of valence or sentiment was exhibited in the true-false assessment: "How am I supposed to know unless I memorize the book?" or "I was confused on what was being asked…" or "I am thinking, 'Boy, I sure do not know this book'" or "I am a lot less trusting of this question now." One senses some negative emotions of frustration and perhaps agitation or anger. Moreover, the necessity for time limits to the questions was established with this quote: "I thought about the book. Thought about how much effort it would take to find." Finally, some expectations were not confirmed. No respondent specifically mentioned Nelson's rules nor any rules related to non-randomness. No comments about the probability or frequency of occurrence were written. Yet, one person thought of patterns and “parts of patterns,” which led to him/her to suggest that all responses were non-random: “…as I literally followed with any type of pattern...any...” And this comment may have expressed Nelson’s rule one, i.e. an observation greater than three standard deviations from the mean: “I observed whether there were clear differences between the pattern of data points at the end and whether *there was a large deviation from the mean that separated a change in the pattern*.”

Some unexpected results did occur, which were used to improve the final instrument. Since the SPC charts, mean, limits and standard deviation lines are color coded to aid in comprehension, the comment “Well, I’m color blind, so there's problem #1” was an obvious oversight and was corrected in the instructions. The colors were also a technical issue among respondents: “Consider using the color red instead of blue or purple for the lines, as the colors appear similar on my screen.” Similar technical issues with the size of the graphs also occurred despite the recommendation of a laptop or desktop computer in the instructions: “Make the lines larger and easier to see.” Furthermore, two additional true-false guessing strategies were uncovered: first, respondents selected true if the word was common in English usage: “‘with’ is a common word used in sentences.” Second, some respondents chose false because the question created suspicion: “I answered false because I no longer trusted the nature of the question,” and this breach of trust would generate false guessing and type II error.

Final survey creation. Based on the results of the pilot surveys, the complete final version was created, including the following eight changes. First, as requested by the IRB, the disclaimer page was modified along with adding a notification that the graphs were in color. Second, the SPC scenario remained the same; however, additional descriptions included the position of the standard deviation lines. For example, the first standard deviation is “next to the centerline." Third, the time limit for the SPC chart evaluation was raised to 40 seconds. Fourth, the SPC chart responses were changed from true or false to random or non-random. Fifth, for the last question which requested the qualitative assessment, the time limit was removed and the open-ended question "As you are viewing the graphical series question, please write down all thoughts, ideas, or internal mental commentary as they occur to you and which lead to your choice of random or non-random" was added. Sixth, the same adjustment was made to the true-false final question, adding the open-ended question "As you are reading the true-false question, please write down all thoughts, ideas, or internal mental commentary as they occur to you and which lead to your choice of true or false." Seventh, the anchors for the bipolar characteristic questions were changed to "small losses with a probability of a large gain" versus "small gains with a probability of a large loss" and "medium losses with a probability of large gains" versus "medium gains with a probability of large losses." And eighth, the request for generational categories was changed to birth year. Appendix B shows the final survey instrument and the associated changes.

Third pilot trial. Once again, to check for any additional oversights of the entire questionnaire, a convenience sample was chosen for the third trial to test the revisions before submission to Qualtrics. Additional programming help from the Qualtrics online community facilitated much clearer graphs by switching from .JEG file types to .PNG. Also, additional HTML text was created, which allowed respondents to enlarge each of the strings to the original file size. Lastly, the qualifier term of "consistent" was added to the characteristic questions concerning types of payoff. Consideration was given between "consistent" payouts, the idea of regularly occurring (dividend), versus "constant" payouts, the idea of unchanging (bond).

“Consistent” was judged to be most appropriate as the scenario focused on stock returns. This change can be viewed in Appendix Table B8.

Qualtrics feedback. The completed survey was sent to Qualtrics, who assigned a project manager and panel projects team who reviewed the survey logic and attributes. They asked about the first characteristics question which probed for preference between “what I avoided” or “what I did.” As a remedy, additional instructions were provided to respondents to clarify that they were choosing not only a preference but also a degree of preference. Finally, during the initial start-up phase of the roll-out, Qualtrics agreed to restrict survey availability to non-mobile devices only, which would allow for maximum SPC chart visibility. However, if response rates declined, the questionnaire would need to be opened to responses using mobile devices to achieve the required sample size.

Final Study

This section describes the Qualtrics soft launch, the data collection process which necessitated voiding a large number of the first data set, the final data set, the qualitative analyses using NVivo software, and the quantitative analyses using MS Excel, SigmaXL, and SMART-PLS software. This section ends with a discussion of results.

Qualtrics soft launch. Once the final changes were completed, Qualtrics invited respondents for the soft launch as an additional quality check for both the data integrity and research expectations. After a short analysis of the data from 38 respondents, the following modifications were undertaken. First, the median response time was 14.5 minutes, which was 11.5 minutes less than expected. As a Qualtrics best practice, exclusion of responses would occur at one-half the median time to complete or 7.25 minutes; however, these responses were saved and could be reviewed if necessary. Second, additional validation limits, from 0 to 99, were added to years of experience. Some respondents entered the year that they had begun work instead of the number of years working, but the average experience was 13 years. Third, the highest job title and professional qualifications questions were randomized to make sure that respondents did not automatically choose the first or second choice (63% were either CPA or CFA). Fourth, as some respondents did not respect the time limits, i.e., leaving the survey and returning, a review would be allowed for click-to-submit times. The largest time gap in the soft launch was 22 minutes. Fifth, Qualtrics mapped all the IP addresses to verify that people actually took the survey from the United States. Sixth, some of the qualitative responses were difficult to understand, but it was agreed that some leeway was necessary due to the nature of the questions. Finally, as noted in the previous pilot studies, negative sentiment was observed in some responses, i.e., frustration and anger.

While a comprehensive analysis was not undertaken, similar expected and unexpected indications arose. The favorable gender ratio of 58% male and 42% female was achieved despite the lack of quotas. To ascertain whether or not respondents used strategies to guess both SPC and true-false questions, a runs, or alternation test was also performed. For the SPC questions, there were six straight line responses, 14 trends bias responses (under-alternation), no oscillation bias responses (over-alternation) at a 99% confidence level. Similarly, for the true-false questions, there were four straight line responses, 11 trend bias responses, no oscillation bias responses at the same confidence level. Therefore, for the SPC chart questions which have definitive rules for determining non-randomness, 53% of respondents provided non-random responses while in the true-false series, the rate was 40%. To verify that randomization of questions had occurred during the original process, the same tests were performed again. Oscillation and trends were found to be random. Moreover, the average SPC percent correct responses were 49% versus 51% for the true-false, not statistically different (*p-value* = 0.891). Similarly, the type I and type II error rates for SPC were 20% and 31%, respectively, while the true-false rates were 32% and 17%, respectively, which was not statistically significant (*p-values* = 0.236 and = 0.164, respectively). Also, first versus second responses were also verified in both categories. The SPC series was 60% first response versus 64% true for the true-false set, and, while not a statistically significant difference (*p-value* = 0.687), it was consistent with patterns first documented almost 100 years ago. Lastly, both psychological scales were verified to understand the composition of respondents who were higher DC, higher PFI, and those that were equal. 87% (33) of respondents showed a DC tendency, 11% (4) with a PFI tendency, and one with equal responses (2% or 1). The PFI proxy estimate used for the original sample size calculation would have suggested one respondent. Four individuals with higher PFI were found in this small sample, which was an optimistic trend.

First data set. When the first final data set was presented, the same data assessment procedures as the soft opening were followed. In Table 10 below, the relationships between answers to the SPC data versus the true-false are shown. As above, both SPC and true-false were answered in no better than random. However, the SPC charts generated more type II errors, i.e., respondents did not see non-random effects while in the true-false, they observed non-random effects that were actually random. Not only were respondents' answers guesses, but the error rates changed with the type of assessment. The smallest significant effect size of 0.3169 was verified to achieve 0.80 power which yielded a sample size of approximately 850. The odds ratio (OR), as described by Sullivan and Feinn (2012), was recommended for binary questions and is the ratio of one metric to another. OR adds an additional dimension to comprehend effect size.

Table 10.
SPC versus True-False Assessments: A Comparison of Attributes

Relationship	SPC	TF	Count	*P-value*	Effect Size	Odds ratio (OR)
Percentage Correct	0.5008	0.4947	1,027	0.7825	0.0122	1.0123
Percentage Type1 Error	0.1784	0.3211	1,027	**0.0000**	**0.3326**	**1.7993**
Percentage Type2 Error	0.3207	0.1842	1,027	**0.0000**	**0.3169**	**1.7412**
First Answer Bias	0.6423	0.5902	1,027	0.0150	0.1073	1.0884

Note. Odds Ratio from Sullivan and Feinn (2012). Significant relationships are shown in bold.

Several other checks were undertaken to review the underlying data.

Following Hair et al. (2017) and Christensen et al. (2014), straight-lining of the psychological scales was observed in approximately 40 instances. Also, reverse coding was explored to check for consistency of responses. The birth years were examined for instances of impossible ages, which were also uncovered. The birth years, when added to years of experience, also created some fantastic dates stretching well into the future (as far as 2040). Clearly, some socially desirable responses were occurring to impress the researcher (Steenkamp, De Jong, & Baumgartner, 2010). Following Sheehan and Pittman (2016) and Wessling, Huber, and Netzer (2017), a simple review of the qualitative responses was undertaken. Based on verbatim text responses, it appeared that one respondent could have taken the instrument more than once, i.e., ballot-box stuffing. Moreover, in the text fields, some respondents typed in random or repeated characters that were gibberish. These concerns were presented to Qualtrics, and their analytics team did a complete review of all the data. Qualtrics reported their results in the email message below, which required the deletion of 472 of the 1,027 total responses:

> "We utilized a combination of available in-platform quality measures and external data cleaning to address response quality. The in-platform features included speeder checks (which is set based on the median soft launch completion time) and duplication checks (via the 'prevent ballot-box stuffing' setting). While we have these in-platform features enabled, we also performed external data cleaning. This process checks the data on the back-end for a variety of behaviors to flag low quality responses -- this includes instances such as straight-liners, geo-IPs outside of the US, gibberish/profane responses, and inattentiveness (high

> survey durations, excessive response selection, etc.). For this particular survey, we also removed responses whose combination of responses seemed illogical (such as starting a career in finance before they were old enough to have a college degree)."

Final data set. When the second final data set was presented, the same procedure was followed as the soft opening and the first final data set. All verifications were made, as previously discussed. The response rate slowed as the second round of data collection began. After two additional weeks had passed, respondents were allowed to use mobile devices. Even though the graph enlargement feature continued to allow excellent visibility, it was also possible that the nature of the responses would change. When the sample reached approximately 1,000 responses, Qualtrics reviewed the data one additional time. About 82 responses were rejected from that set. Then within a couple of weeks, the final sample size of 1,050 was achieved.

In Table 11, the relationships between answers to the SPC chart data versus the true-false assessment are given for the final dataset. The elimination of the questionable responses had also resulted in a reduction in the effect sizes between type I and type II error rates. Yet when examining these effect sizes, power was maintained with a sample size of 850. Moreover, there may have been a device effect. While the sample size for mobile does not meet the correct power level, respondents may be making different selections depending on how they viewed the survey, i.e. more random when looking at a complex graph on a smaller screen.

Table 11.
SPC versus True-False Assessments: a Comparison of Attributes from Final Data

Relationship	SPC	TF	Count	*P-value*	Effect Size	Odds ratio (OR)
Percentage Correct	0.5022	0.4997	1,050	0.9062	0.0051	1.0051
Percentage Type1 Error	0.2119	0.2947	1,050	**0.0000**	**0.1911**	**1.3912**
Percentage Type2 Error	0.2859	0.2118	1,050	**0.0001**	**0.1717**	**1.3496**
First Answer Bias	0.5740	0.5397	1,050	0.1126	0.0692	1.0637
Percentage Correct (non-mobile device)	0.5019	0.5001	899	0.9380	0.0037	1.0037
Percentage Type1 Error (non-mobile device)	0.2046	0.2756	899	**0.0004**	**0.1667**	**1.3472**
Percentage Type2 Error (non-mobile device)	0.2935	0.2244	899	**0.0008**	**0.1583**	**1.3084**
Percentage Correct (mobile device)	0.5043	0.4974	151	0.9038	0.0139	1.0140
Percentage Type1 Error (mobile device)	0.2553	0.2368	151	0.7083	0.0431	1.0783
Percentage Type2 Error (mobile device)	0.2404	0.2368	151	0.9408	0.0085	1.0154

Note. Odds Ratio from Sullivan and Feinn, (2012). Significant relationships are shown in bold.

Given the device visibility issue and consistent with Blair and Zinkhan (2006), an assessment of non-Response bias was made between the first and last halves of the samples for both the SPC and the true-false assessments. Table 12 outlines the results with the SPC instrument. No *p-values* were significant, which indicated respondents answered similarly regardless of when they viewed the survey.

Table 12.
SPC Non-Response Bias Assessment: Front Half versus Back Half Sample

Relationship	SPC	SPC	Count	*P-value*	Effect Size	Odds ratio (OR)
Percentage Correct	0.5045	0.5000	525	0.8848	0.0090	1.0090
Percentage Type1 Error	0.2056	0.2181	525	0.6211	0.0305	1.0607
Percentage Type2 Error	0.2899	0.2819	525	0.7744	0.0177	1.0284

Note. Odds Ratio from Sullivan and Feinn, (2012). No significant relationships were found.

A similar view of non-Response bias for the true-false assessment is visible in Table 13. As with the SPC test, no *p-values* were significant; therefore, no evidence of non-Response bias. Lastly, the open-ended questions were reviewed to determine possible duplicate respondents. Since none were found, the final data set was approved.

Table 13.
True False Non-Response Bias Assessment: Front Half versus Back Half Sample

Relationship	TF	TF	Count	*P-value*	Effect Size	Odds ratio (OR)
Percentage Correct	0.5008	0.4986	525	0.9435	0.0044	1.0044
Percentage Type1 Error	0.2652	0.2747	525	0.7245	0.0218	1.0364
Percentage Type2 Error	0.2340	0.2267	525	0.7780	0.0174	1.0324

Note. Odds Ratio from Sullivan and Feinn, (2012). No significant relationships were found,

Descriptive statistics of the final data set. The final sample of 1,050 responses was composed of 64% male and 36% female. Using generational categories as proposed by Dimock (2019), respondents were 54% Millennial, 35% Generation X, 11% Baby Boomers, and one respondent was from the Silent Generation. Fifty-two percent of the individuals had some college or a college degree, while 47% had some graduate work or a graduate degree. Seventy-four percent of respondents' principle work effort was quantitative, 46% were affected by employment in a financially distressed company, and 69% suggested life stress might affect their business perspective. The minimum experience required to complete the survey was five years. One respondent had 48 years of experience while the average was 13 years, and all were based in the United States.

Qualitative results. This step was originally proposed to be the last one but was repositioned after dataset approval to avoid any bias that might be created from the quantitative analysis. As previously noted, Nickerson (2002) suggested that respondents should be allowed to explicate their knowledge of randomness; therefore, one question from each of the SPC and true-false sections asked them to write aloud. The analysis for both the SPC and true-false assessments followed Creswell (2009). Open codes were developed for the SPC graph using concepts from literature and the results of the pilots. For the former, when respondents were searching for non-randomness, several concepts emerged: Clustering, data points falling on one side of the median (next to each other); Mixture, an absence of data points near centerline;

Oscillation, data points constantly fluctuating up and down; and Trends, where seven or more data points continuously increased or decreased. These types of comments occurred in about 22% of the commentaries. For example, respondents described oscillation as “I looked to see if there was a pattern between where the up and down occurred,” or “There are various ups and downs, and the middle line is not consistent, so I have chosen randomly.” As an example of mixture, one respondent noted, “Fairly consistent trading pattern first two weeks and the three days of volatility before returning to its norm.” This was the correct assessment of the graph, yet only 4% of respondents made the correct observation. About 8% used some form of incorrect logic. As an example of this, “Random is due to the two big falls at the end of the graph,” which was correct but drawing an incorrect conclusion.

Other open codes included valence, both positive and negative. For positive valence, the sentiment was, as Steenkamp et al. (2010) discussed, socially desirable responding but aimed at the researcher: “This graphical series question is very good.” The negative valence examples were respondents who did not understand what was being measured, were frustrated, and sometimes angry: “This is a very odd survey, I don’t know what you are getting at?” or “I don’t see the point of this exercise. Can you tell me the point of this exercise?” Valence represented about 12% of the responses, 10% positive and 2% negative.

Respondents also discussed patterns and the absence of patterns about 10% of the time: “I did not see many patterns. Most were random,” or “There is no repeating pattern in this sequence.” But also, quite the opposite view was present: “It is very much random because of the way the pattern goes.” In this instance, randomness has a pattern that conveys two opposite views. Similarly, respondents had personal theories and described them, “A random walk of length on a possibly infinite graph with root is a stochastic process with random variables.” This fact may be true from a theoretical perspective; however, the observations contained only 18 data points. Another example included “It's non-random because points 2, 10, 17, 18 are on the positive side while points 4, 7, 12, 13, 14, 15, 16 are on the negative side. Meanwhile, the number of points in the line of alignment is lesser than the number of points in disarray.” This philosophizing occurred about 12% or the time.

Finally, about 25% did not comprehend the exercise, repeated the question, or otherwise guessed: “I just see and guess them. I base it on the image and those values to decide it's random or non-random,” or “The randomness of individuals sometimes tends to decline for no apparent reason.” The inability to articulate an answer will be examined in the discussion section below and Chapter 5. Figure 5 shows a word cloud based on respondents who guessed, described, or repeated the question using the synonym setting and most frequent 1,000 words with five or more occurrences.

Figure 5. Word cloud of SPC Open-ended Question. The actual answer was non-random. These are the words associated with a random response.

Other random/non-random keywords mentioned by respondents included: predictability, symmetry, normal, correlations, stable, disorder, luck, repetition, sporadic, parallel, consistent, reliable, chaos/volatile, slope, absolute, control, rationality, certainty, and outlier, but these were rare. Also, what was not present was anyone citing Nelson's rules or the 68-95-99 standard deviation rule for a normal distribution, a fundamental statistical concept.

The open coding for the true-false assessment was much simpler. Positive and negative valences were 19% and 15%, respectively, or 34% of the total: "I like it, is innovative and fun, stirs my emotions," versus "May I confess? This exercise makes absolutely no sense to me, and I don't know what information you're expecting to glean from surveying people about the placement of words in various texts." Many respondents assumed that they could guess better through a personal theory based on English language usage or equated the question to "True, False, Not Given" responses to reading comprehension exams. This occurred about 14% of the time, but the majority, 51%, just guessed, described, or repeated the question. Figure 6 shows the results of the guess category following the same criteria used for the SPC instrument. In comparing the clouds, two of the most frequents words, "random" and "false," are both incorrect answers. These responses are narratives and do not reflect any real understanding of probability or the rules related to non-random strings. This comparison further confirms the idea that guessing occurred widely in both instruments and which converges with the quantitative assessment of percentages correct for both tests.

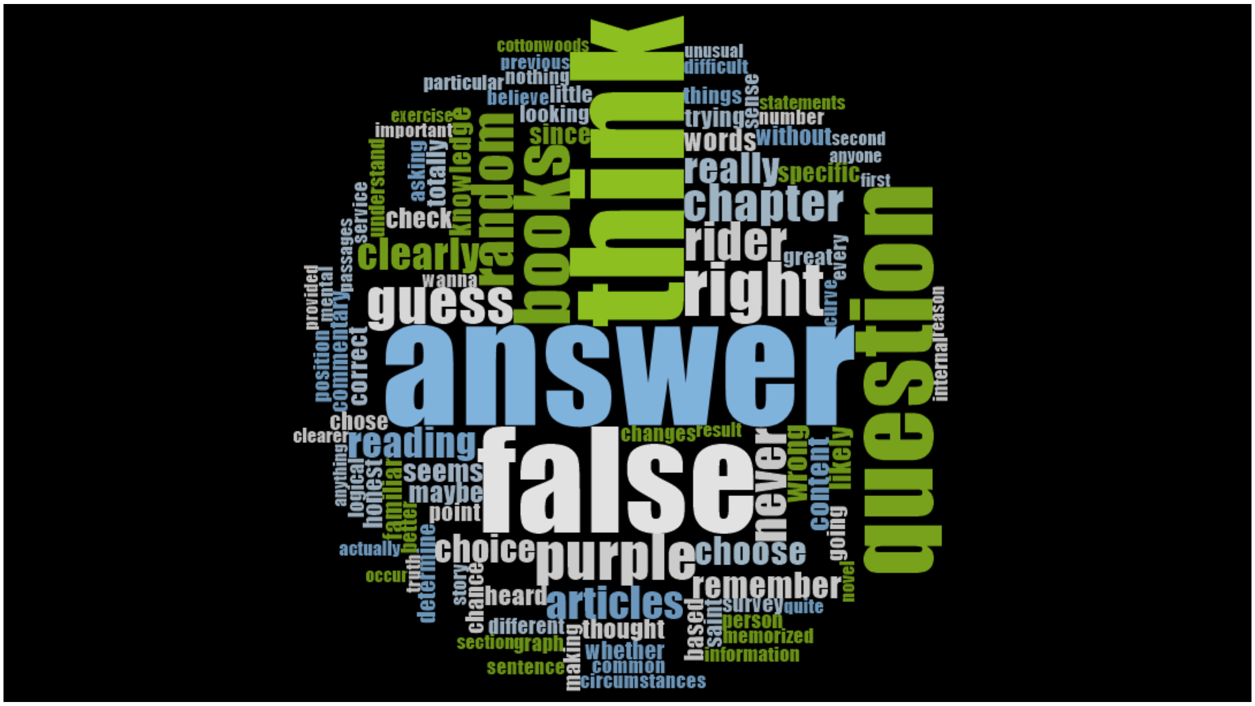

Figure 6. Word Cloud of True-False Open-ended Question. The actual answer was true. These are the words associated with a random response.

Creation of Cohen's kappa grouping variable. The expectation that the data set would contain high alternation bias (oscillation) was not met. In initial simulations of 1,000 data points, the result should have been approximately 400 unique strings with groups of ten or larger ranging from 10 to over 70 respondents, e.g., those TF and FT repeated ten times. However, in the data from actual respondents, there were 833 individual groups with only three groups with ten or more data points. Moreover, the Kappa calculation did not allow for values that could be assembled into groups with the use of a caliper technique (a range of Kappa values).

In the SPC assessment, respondents returned 591 non-random patterns, which were calculated using the runs test for oscillation and trend within the SigmaXL software. Over-alternation was expected to be the dominant pattern but occurred only 22 times. Straight-lining first and second responses occurred 103 and 33 times, respectively, while under-alternation occurred 433 times. A similar pattern was repeated in the true-false assessment: over-alternation

37, either all responses of true or false, 65 and 82, respectively, and under-alternation 368 times for a total of 552. Once again, these results showed how respondents turned a fixed randomized order into patterned responses.

To overcome the lack of oscillation bias and continue the idea of groups based on the manner that questions were answered, the true-false non-random respondents were organized as follows: All true (65 elements), "more true" under-alternation (174), more false under-alternation (81), all false (82), over-alternation (37), and the balance of the sample (611). Figure 7 shows the kappa values of group two or the "more-true" under-alternation group. As shown, grouping towards true or false tendencies created an unusable wide range of Kappa values.

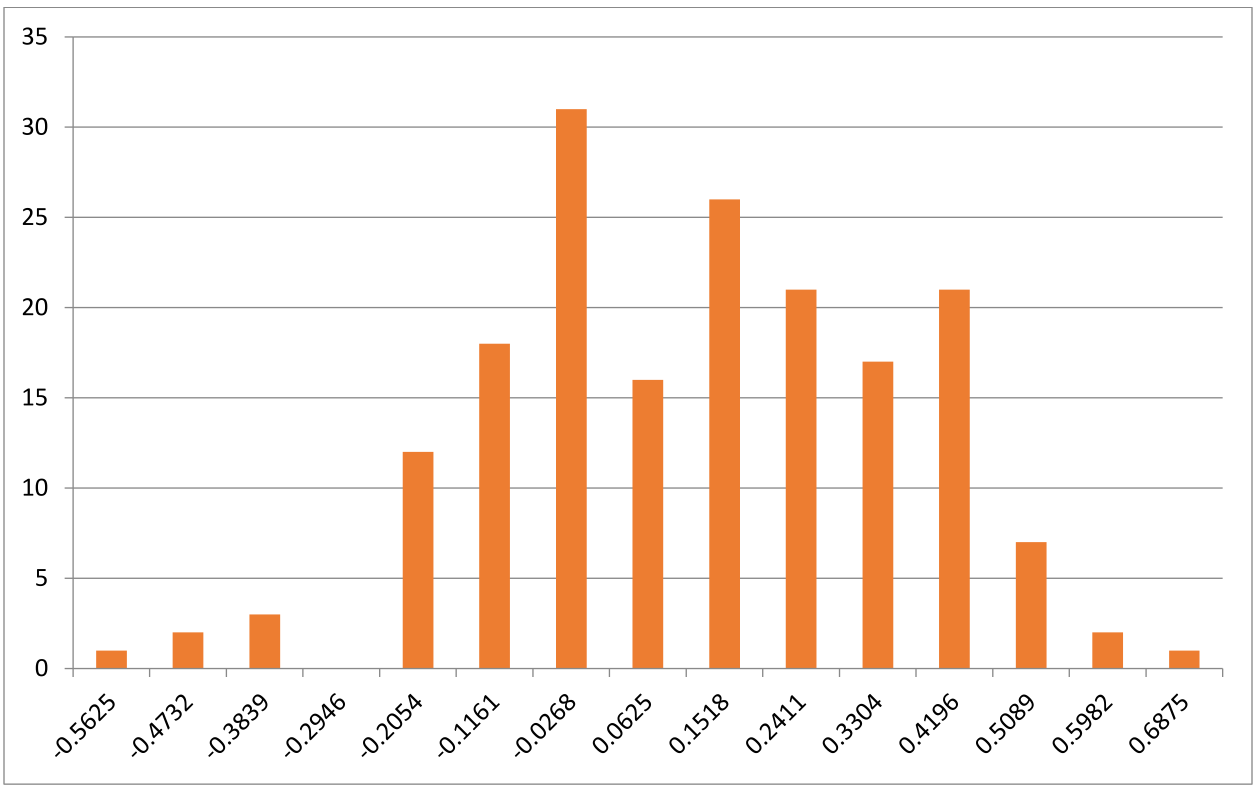

Figure 7. *More True Under-alternation Group Cohen's Kappa Values, Group Two.*

SEM measurement model. Following Klarner, Sarstedt, Hoeck, and Ringle (2013) and Hair et al. (2017), the outer models with both reflective and formative constructs were reviewed.

All constructs were assessed for reliability and validity. Desirability of control (DC) and personal fear of invalidity (PFI), the two reflective measures, were analyzed first by examining convergent validity through their outer loadings which are visible in Appendix Table C1. Values greater than 0.7 were acceptable, while those between 0.4 and 0.7 needed additional evaluation. DC7, DC10, DC16, DC19, DC20, were accepted while PFI3, PFI4, PFI5, PFI6, PFI13, and PFI14 also passed.

Next, the formative measures of the characteristics of type I and type II errors were examined for convergent validity. First, an evaluation of the significance of the outer weights was completed with results visible in Appendix Table C2. Second, if the characteristic was not significant, the value on the outer loading was reviewed. Those above 0.5 were retained, while those below 0.5 were excluded. C2 and C2R were retained for significance, while C3, C3R, and C4 were retained based on their outer loadings. In Appendix Tables C3 and C4, the improvements in convergent validity are shown by the increase in the average variance extracted (AVE).

Similarly, the improvements to the reliability measures of Cronbach's alpha and composite reliability can also be seen in Appendix Tables C3 and C4. The SPC error rate reliability measures are 1.00 by definition of the errors themselves, missing an effect or type II error versus observing an effect that is non-existent or a type I error. Next, the variance inflation factors (VIF) were examined for collinearity among the items, and DC and PFI passed. The characteristics of type I and type II errors had no values because of the mirror image effect employed in this study. And finally, to measure discriminant validity, the Fornell Larker criterion was used for the reflective measures, DC and PFI, which are visible in Appendix Tables C5 and C6. The highest row values for DC and PFI were achieved. For single measures, such as type I and type II error rates and formative measures, Hetero-Trait Mono-Trait (HTMT) values were evaluated and are listed in Appendix Tables C7 and C8. A complete bootstrap procedure was run to see if the 95% confidence intervals contained the value 1.0. All measures passed this discriminant validity verification; therefore, based on the results of the updated model, all measures of reliability and validity were achieved. The next step was to evaluate the structural model and relationships with the latent constructs.

SEM structural model. A standardized procedure was followed to analyze the structural model and the proposed hypotheses that were investigated. In Figure 8, the structural model, path coefficients, and R^2 are shown.

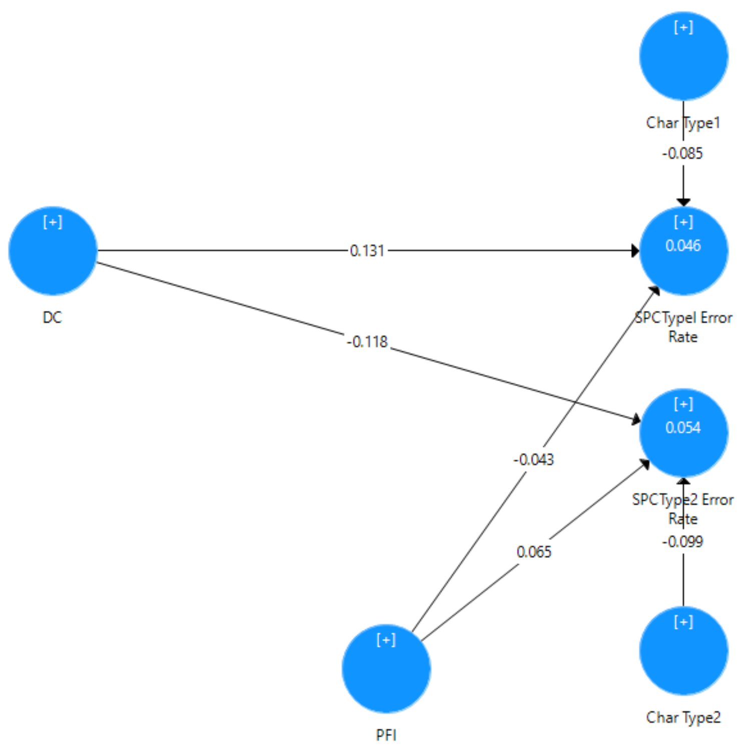

Figure 8. *Structural Model with Path Coefficients and Respective R^2.*

Specifically, R^2 for type I and type II errors rates were 0.046 and 0.054, respectively, and R^2 adjusted were 0.043 and 0.051, respectively. F^2, which is an evaluation of effect size, is shown in Table 13 below. Those values indicate, however, an extremely weak relationship because a typically weak result carries a value of 0.02; the values in Table 13 are between 2.5 and 20 times inferior to a weak result. In Appendix Table C9, the structural models' path coefficients and significance levels are listed. Hypothesis H1, stating DC provides a positive path to the type I error rate, was supported (mean = 0.131, standard deviation = 0.048, *p-value* = 0.007). Therefore, higher DC individuals would generated more type I errors. Also, as expected and not hypothesized, DC had a significant negative relationship with the type II error rate. Hypothesis H2, stating PFI provides a positive path to the type II error rates was not supported (mean = 0.065, standard deviation = 0.051, *p-value* = 0.198). Therefore, higher PFI individuals would not necessarily generate more type II errors. PFI also had a non-significant negative relationship with the type I error rate.

Table 14.
F^2 Evaluation of Effect Sizes

Measures	SPCType2 Error Rate	SPCType1 Error Rate
Char Type1		0.006
Char Type2	0.008	
DC	0.006	0.007
PFI	0.002	0.001

Note. Values for f^2 *for small, medium or large effects are 0.02, 0.15, and 0.35, respectively.*

The characteristics for a type I error rate were C2, C3, and C4 which were supposed to have a significant positive relationship with type I error rates, but, in fact, were negative (mean = -0.085, standard deviation = 0.032, *p-value* = 0.009); therefore, H3 was not supported. Therefore, gains to losses did not generate more type I errors. Similarly, the expectation that the reverse characteristics, C2R and C3R, would provide a positive path to type II errors, H4, was not supported and negatively associated (mean = -0.099, standard deviation = 0.033, *p-value* = 0.003). So, losses to gains did not generate more type II errors. The reason for the directional switch between types I and II errors was related to question design versus respondents' interpretation. Respondents interpreted the payoff curves in a conventional way: win-stay (no switching which yields type II errors), lose-shift (switching which yields type I errors). Yet, the question was designed to differentiate between both conventional and contrarian views, i.e., a win-stay, lose-shift strategy as opposed to a win-shift, lose-stay strategy (see Scheibehenne, Wilke, & Todd, 2011).

To assess the predictive value of the model, Q^2 was calculated using a blindfolding procedure. A default distance of 8 was selected based on a software technical requirement that the sample size divided by the distance needed to be a non-integer result. Q^2 values exceeding zero are considered predictive. Type I error rate and type II error rate predictive values were 0.039 and 0.050, respectively. The calculated Q^2 effect sizes were -0.011 and 0.012, respectively, and extremely weak.

Hair et al. (2017) explicated model fit measures within Smart PLS, which included goodness-of-fit (GoF), standardized root mean square residual (SRMR), and root mean square residual covariance (RMS_{theta}). The authors recommended not to use any of these measures as the current research did not have solid positive results. Some promising studies pointed to RMS_{theta} as a promising metric, so, for completeness, the value in this study was 0.20.

Multigroup analysis. Following Hair, Sarstedt, Ringle, and Gudergan (2018), there are several methods to test groups within Smart PLS. The measurement invariance of composite models (MICOM) necessitates three qualifying conditions to calculate differences between two groups. These conditions are configural invariance, compositional invariance, and equality of composite mean values and variance. Another technique available in Smart PLS is the Multi-Group Analysis (MGA) procedure, which applies a permutation calculation.

To use either of these techniques for the six Kappa groups, one group would need to serve as an anchor, which would logically be the Kappa/Alternation group six containing the responses with no oscillation or trend bias. Group six would be the anchor to which other groups will be compared to avoid a familywise error. Appendix Table C10 lists the path coefficients and *p-value*s for Kappa group six. Unfortunately, this method could not be employed as the group six model had values that were different from the complete model; therefore, each group model was assessed individually. It was possible to retain configural invariance from the MICOM criteria. Hair et al. (2018) stated that for configural invariance to exist identical indicators, identical data treatment, and identical algorithm usage must occur. Since the respondents were grouped from the same study, and the exact measurement model was used along with identical optimization techniques, the measurement model passed configural invariance criteria.

Furthermore, Hair et al. (2018) recommended that all groups should meet a minimum sample size, which was determined by the maximum number of arrows pointing to a latent construct multiplied by ten. For this study, a minimum of 30 was needed for group formation. For the five groups to be tested, all met minimum sample size requirements.

The two hypotheses to be assessed based on the Kappa/Alternation groups were as follows: hypothesis H4 with a positive path to type I error rates using groups one and two (straight line true, true/under-alternation); hypothesis H5 with a positive path to type II error rates using groups three and four (false/under-alternation, straight-line false). Hypothesis H10, where true responses exceeded false responses leads to a positive path to the type I error rate, was also scrutinized.

Appendix Tables C11 and C12 show the path coefficients and *p-values* for Kappa Groups one and two: those that straight-lined all true and those that had true tendencies with under-alternation bias. In the DC to type I error rate relationship, group one results were mean = 0.203, standard deviation = 0.199, and *p-value* = 0.308, while group two results were mean = 0.113 standard deviation = 0.134, and *p-value* = 0.399. Therefore, hypothesis H5, a group with a positive significant path to type I error rate, was not supported, i.e., the tendency to respond true did not generate type I errors.

In Appendix Tables, C13 and C14 are the path coefficients and *p-value*s for Kappa Groups three and four: those that straight-lined all false and those that had false tendencies with under-alternation bias. In the PFI to type II error rate relationship, group three results were mean = 0.124, standard deviation = 0.182, and *p-value* = 0.495, while group four results were mean = -0.233 standard deviation = 0.224, and *p-value* = 0.299 (not in the correct direction). Therefore, hypothesis H6, a group with a positive significant path to type II error rate, was not supported, i.e., the tendency to respond false did not generate type II errors.

Appendix Table C15 presents the path coefficients and *p-values* for higher percentage true responses from the true-false assessment. In the DC to type I error rate relationship, this group's results were mean = 0.129, standard deviation = 0.073, and *p-value* = 0.076. Despite being close, hypothesis H10, with a positive significant path to type I error rate, was not supported, i.e., the tendency to respond true did not generate type I errors.

Analysis of final three hypotheses of this study. A major expectation in this study was the presence of alternation bias. Hypothesis H7 predicted that the alternation rate in the true-false assessment would be significantly higher than a random sample. As noted in Chapter 3, Griffiths et al. (2008) derived the proof of the probability calculation for cases of three to 14 strings, which yielded an exponential function of $2.5593e^{-0.452}$. Therefore, the probability of 20 randomly alternated true-false responses is 0.0003 or 3 basis points. Furthermore, a simulation was created using the binomial inverse function in Microsoft Excel with parameters of 1 trial, 0.50 probability, and the random number generator repeated 1,050 times. The same SigmaXL runs tests, the above/below non-parametric exact runs tests, were completed with *p-values* at a 99% confidence level. The results of these simulations showed no alternation bias or straight-line bias at all. However, there was excessive trend bias in 247 cases or 24% of the sample. In the true-false assessment, there were 37 observations with oscillation bias. Even with only 37 observations, the difference was significant in a t-test (*p-value* = 0.000, effect size = 0.378). In fact, to be significant at the 95% confidence level, only four observations were required. Therefore, hypothesis H7 was supported but not in an expected manner. The expectation of high oscillation bias was not achieved as expected in the literature.

Hypothesis H8 predicted that the percentages of correct responses would be significantly higher in the SPC assessment than in the true-false assessment simply because finance and accounting personnel would know and recognize non-randomness, would know Nelson's rules, or would have ideas about outliers, trends, oscillation rates, etc. From Table 11, it is clear that respondents guessed or had incorrect strategies as there was no significant difference between the two assessments. (*p-value* = 0.9062, effect size = 0.0051, and odds ratio = 1.0051). Therefore, H8 was not supported. This lack of random sequence knowledge in corporate finance and accounting personnel was disappointing but not surprising.

The final hypothesis to be evaluated was H9, which predicted that the percentage of true responses in the true-false assessment group would be significantly higher than the percentage true in a random distribution, which, based on the discussions in Chapter 3, has a 100-year tradition. A t-test of the percentage true rate of 0.5397 against 0.500, a random expectation, produced a *p-value* = 0.069, effect size = 0.070, and odds ratio = 1.079. Running 10 simulations with sample sizes of 1,050 observations, for a total of 10,500 using the inverse binomial function, as previously described, yielded the results of mean = 0.5010, standard deviation = 0.0037, skewness = 0.0071, and kurtosis = -1.4675. The t-test calculation resulted only in miniscule changes from 0.5000. Examining the confidence intervals using the upper limit 0.512, yielded a non-significant *p-value* = 0.205, effect size = 0.055, and odds ratio = 1.054. Using the lower limit of 0.490 produced a significant *p-value* = 0.023, effect size = 0.099, and odds ratio = 1.101, however, in the literature, the expectation was closer to 0.60 true, i.e., when faced with a random choice individuals guess true significantly more often than false. Therefore, hypothesis H9 was not supported.

Modeling unobserved heterogeneity. Since oscillation, under-alternation, and Kappa groupings did not find any significant paths. The next step was to test for unobserved groups within the dataset. Hair et al. (2018) indicated that traditional clustering methods such as k-means clustering did not meet expectations for a PLS model. Instead, the authors recommended methods such as latent class techniques or response-based segmentation techniques. Furthermore, they offered a step-by-step method to investigate unobserved heterogeneity. First, finite mixture partial least squares (FIMIX-PLS) calculated several fit indexes for any number of segments based on theory or research preference. Segments that contained the lowest calculated index value were considered optimal choices. SMART-PLS contains ten different indexes. Some of those types include Akaike's Information Criterion (AIC), Modified Akaike's Information Criterion with Factor 3 (AIC_3), Modified Akaike's Information Criterion with Factor 4 (AIC_4), Bayesian Information Criterion (BIC), Consistent Akaike's Information Criterion

(CAIC), and Minimum Description Length 5 (MDL_5), among others. Second, the prediction-oriented segmentation partial least squares (POS-PLS) was executed, which moved data points between the selected number of groups to maximize explained variance. Once these groupings were completed, it was the responsibility of the researcher to explain the structure, if any, and to re-examine group-specific models using the PLS-SEM procedures previously described

In Appendix Table C16, 11 separate fit indexes are listed across eight segments that were calculated in eight separate trials. The minimum sample size criteria was also based on the largest number of arrows directed to any single latent variable divided by the sample size; however, 35 groups were not feasible. Assuming group sizes of between 60 and 210 respondents or roughly 5% to 10% of the sample leaving a large unclassified group would seem to follow leading research on personality types (see Gerlach, Farb, Revelle, & Amaral, 2018). Therefore, the number of groups might be between five and seven. Furthermore, some index combinations provided more accurate estimates of the explained variance than others. According to Hair et al. (2018), the pecking order of matching indexes to estimate segments was as follows: AIC_3 and CAIC, AIC_3 and BIC, and AIC_4 and BIC. The latter combination, whose actual minimized values are shown in Appendix Table C16, indicated the possibility of six segments. Also, the entropy statistic normed (EN) of 0.643 was larger than 0.500, the hurdle rate, and provided additional confirmatory evidence of six groups.

Next, each segment group was reviewed for the minimum size relative to the sample of 1,050 respondents. Appendix Table C17 indicates each of the segment sample sizes met the 30-respondent minimum. And finally, Appendix Table C18 shows the estimated R^2 values of each segment. Segments three to six have R^2 values over 80%, while segment two is approximately 45% in opposition to the 5% rate of the overall model.

With the FIMIX-PLS segment solution determined, the next step was to execute the prediction-oriented segmentation (POS-PLS) procedure. By random assignment and using an optimization procedure that was designed to maximize the explained variance of the endogenous variables, respondents were placed in one of the six groups (Hair et al., 2018). Appendix Table C19 lists the path coefficients for each of the POS groups, and Table C19 indicates the R^2 values. Once each respondent had been assigned to a group, a determination was made as to the characteristics of these groups. Initially, a review of the qualitative responses was undertaken to see if patterned commentary or specific coding had occurred among segment members. Subsequently, a review of the quantitative data, including the screening criteria and several of the demographic variables, was considered. Unfortunately, no information was forthcoming. However, several of the segments revealed a type of finance/accounting personality based on an approach to errors which was much richer than anticipated.

Specifically referring the Appendix Tables C19 and C20, segment six indicates either an individual with high DC or one with high PFI with strong positive tendencies to the type I error rate and a strong negative tendency to the type II error rate. These individuals could be described either as “control freaks” or perhaps micromanagers or as super fearful persons who check everything multiple times. These activities would cause many type I errors through the effort to verify nonexistent effects. These could be the individuals most subject to maladaptive behaviors described by Fenton-O’Creevy et al. (2003). In total opposition to what was expected, segment five individuals had stronger tendencies of DC to type II error rates and PFI to type I error rates. These controlling individuals perhaps know accounting and finance subject matter and may be exhibiting some hubris or other IoCON behavior, which leads them away from seeing any effects (type II).

Conversely, the semi-fearful finance individuals will be checking for effects constantly and, therefore, making more type I errors as in segment six. And lastly, segment two individuals are similar to those anticipated in this study. These individuals have a weaker tendency of DC to the type I error, but it remains significant. They control and make type I errors, but much less than segment six. Moreover, segment two also contains the most fearful finance individuals of all segments, the highest path of PFI to type II error rate. These individuals are so fearful that this behavior leads them to indecision and IoCHA. Their remedy would be to follow the advice of Harris and Osman (2012) and "pretend" they had control despite their perception of chaos.

Once the groups were identified, it was possible to add the POS-PLS groups to the data and calculate additional individual segment models. As an example of this, Appendix Table C21 shows segment six's path models and *p-values*. All paths are significant except for the characteristics type I to the type II error rates (compared to the original model in Appendix Table C9). For the type I error rate, the R^2 and R^2 adjusted were 0.983 and 0.982, respectively, while the type II error rates were 0.986 and 0.985, respectively. Moreover, when comparing the effect sizes of DC and PFI, f^2, listed in Table 14, to those below in Table 15, these were extreme values.

Table 15.

F^2 Evaluation of Effect Sizes for POS Segment Six (an example)

Measures	SPCType2 Error Rate	SPCType1 Error Rate
Char Type1		0.032
Char Type2	0.216	
DC	47.382	39.440
PFI	0.032	9.067

Note. Values for *f^2 for small, medium or large effects are 0.02, 0.15, and 0.35, respectively.*

True-False SEM PLS model. Rather than knowing the rules of randomness within time-series variation, respondents basically guessed answers not only to the true-false assessment but also to the SPC assessment as repeatedly stated. Therefore, there was another possible PLS structural model that was created using the true-false assessment as endogenous variables. For parsimony and the fact that this model was not anticipated, only the model and final results were presented. The question to answer was whether or not respondents answering both assessments would group similarly in each segment.

In the same standardized procedure as the SPC PLS model, the true-false model was constructed. First, convergent validity was assessed for the measurement model. Outer loadings were examined for the reflective variables of DC and PFI, and outer weights and loadings for the formative characteristic variables. Figure 9 shows the current model. DC2, DC4, DC5, DC11, DC12, and DC15 were retained for the desirability of control. PFI3, PFI4, PFI5, PFI6, PFI7, PFI 11, PFI12, and PFI14 were retained for personal fear of invalidity. Characteristics C2, C5, C2R, and C5R were also kept. The structural model in Figure 8 below was reviewed for reliability (Cronbach's alpha, composite reliability, and AVE), discriminant validity (Fornell-Larker and HTMT), collinearity (VIF), and goodness of fit with the same caveats.

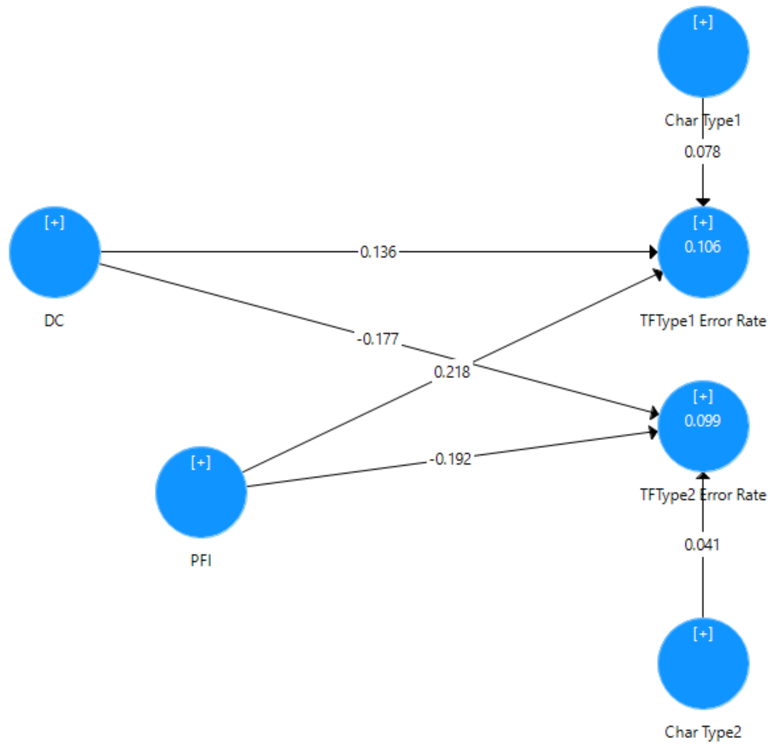

Figure 9. *True-False Structural Model with Path Coefficients and Respective R^2.*

Once the structural model was completed and path coefficients were examined, along with their significance, through the bootstrapping technique, the blindfolding procedure was used to review out of sample predictive power. Next, the FIMIX-PLS process was used to determine the number of segments, shown in Appendix Table C22. The segment sizes prohibited any segments over six, but since no fit indexes were minimized at that number, the optimal choice of five segments was selected.

The POS-PLS process was completed to assign each respondent to one of the five groups. In Appendix Table C23, R^2 are listed for each segment, as are the weighted average and original values. Overall, R^2 remained weak, yet some of the segments had high explained variance though not as elevated as the SPC model. The path coefficients are listed in Appendix Table C24. The hypothesized paths of DC to type I and PFI to type II are seen in segment five, which is similar to SPC segment two. The true-false segment two showed a strong relationship between PFI and type I error rates as in SPC segment five but without a strong DC to type I error rate relationship. A crosstab count of the respondents in both the SPC and true-false assessments is shown in Appendix Table C25. While the initial reaction to this difference might appear problematic, Table 11 showed that the same individuals reacted in different ways to each assessment, i.e., missing effects in the SPC graphs, and perhaps, overthinking in the true-false guessing game. Moreover, this may provide evidence for Brunswik's representative design, which is discussed in the next section (see Gigerenzer, 2000).

Discussion

The most interesting finding and contribution of this study is that accounting and finance personnel, when faced with time series, do not appear to be able to distinguish between a random series and a non-random one. Our findings clearly showed that where the time-series graph had no positive or negative slope overall, the results of the individual assessments were no better than guessing. Likewise, this fact was evident in the qualitative responses. Respondents were more likely to admit to guessing in the true-false assessment, but roughly the same percentages (12%-13%) had developed their own personal logic, narrative, or rationalization. While some respondents did know what to look for statistically, i.e., oscillations, trends, mixture, stratification, etc., actually operationalizing this general knowledge into correct answers was problematic. And no apparent knowledge appeared to exist concerning Nelson's rules or the 65-95-99 rule of standard deviations in a normal distribution. As over half had college degrees and a large portion, at minimum, some graduate work, one might anticipate some exposure to a statistics class that was perhaps long forgotten but urgently required.

The second finding is the obvious observation of the inability of respondents to write down their thought processes coherently. They would easily move off topic, repeat the question, provide complementary phrases, or even turn to frustration and anger. However messy this process was, it fulfilled the idea suggested by Nickerson (2002) that subjects should be able to write about what they know about randomness. Microphenomenology, as a qualitative research technique, may provide a solution to this issue and will be discussed in Chapter 5.

Third, the assessments themselves revealed something about human adaptability. Table 11 showed that in the SPC assessment, there was a significance in type II errors, i.e., missing the non-random effects within the series. However, in the true-false assessment, the tendency moved to type I errors, i.e. seeing effects that were not there. What was designed to give purely random answers was instead used by some who attempted to "see" what the researcher was up to or to ascertain some hidden knowledge. The same respondents, in effect, guessed one way in the SPC assessment and another in the true-false. Gigerenzer (2000) described Egon Brunswik's probabilistic functionalism which was composed of four precepts: achievement when individuals attempt to adapt to their environment; the ambiguity of cues that exist in that environment; vicarious functioning when those individuals attempt to process environmental cues by recombination and re-substitution; and representative design which researchers should use both to capture this vicarious functioning and to generalize beyond just one assessment of stimulus/response. In fact, this study also measured randomness in two ways (complex versus simple) which provided two error tendencies (missing effects, seeing non-existent ones) and which points to the insight of Brunswik's argument, i.e., without both assessments, those insights would have gone undetected.

Fourth, the SEM-PLS modeling process was a powerful tool to understand latent construct relationships. However, the models that were created revealed minimal explained variance and were unusable for any predictive assessment. While directionally correct, the path between the desirability of control and the type I error rate was the only significant one. And the characteristics, while showing significant paths, were inverse, e.g., gains to losses were associated with a type II error rate. Therefore, as a generic assessment tool to identify illusion of control or illusion of chaos, SPC and true-false methods were inadequate. And yet, the quest for an objective error rate instrument that measures all individuals in a sample as hoped for by Fenton-O'Creevy et al. (2003) must continue.

Fifth, Smart PLS multi-group analysis would have been effective with only two groups. However, the FIMIX-PLS and POS-PLS represented newer tools to maximize explained variance. These tools uncovered more complex relationships between control, fear, and error rates. The use of these modules to create groups from which individuals could then be interviewed and from whom characteristics could be determined marked a path forward in the studies of illusion of control and illusion of chaos. Moreover, using new statistical methods with older constructs coincides with the "tools to theories" heuristic advocated by Gigerenzer (1991, 2000), i.e., new tools provide undiscovered insights into older ideas or theories.

Sixth, the use of Cohen's kappa as a grouping variable was original but depended heavily on the presence of over-alternation bias. Without this repeated oscillation, the respondents' samples would not fall into self-forming groups. Likewise, under alternation did not indicate a particular style or behavior to be used as a grouping variable. So as a practical grouping tool, kappa has its flaws regardless of the alternation rate.

And, even though this particular model was not performing as expected, finance and accounting personnel should reflect on the potential types of errors they might make when executing all aspects of their craft: from accrual estimates to financial forecasts and analyses. Remembering Taleb (2014), providing random forecasts increased risk profiles, and Stimmler

(2013), increased complexity without new information increased risk profiles, may help finance professionals not to go the way of the specialist, Lem Put, privy builder (see Sale, 1929).

As a final note concerning data integrity in this study, Qualtrics data sets, PLS-SEM models and data uploads will be available to interested parties upon request. The reporting of the analytics of this study were those items actually used for evaluation and were typically reported in previous studies.

CHAPTER 5 – CONCLUSIONS, LIMITATIONS, AND FUTURE RESEARCH

Conclusions

Several conclusions from this study are evident. First, statistical process control charts and true-false guessing methods do not explain much variance in type I and II error rates for all individuals. Second, respondents were given a voice to explain, however awkwardly, their thoughts on randomness. Initially, there was some disappointment in the quality of the responses; however, this sentiment changed with the realization that respondents may not have this capability. Third, the use of advanced structural equation modeling, SEM-PLS, FIMIX-PLS, and POS-PLS, offer new exploratory tools of study. Fourth, finance and accounting must have general knowledge about randomness, variation, and error implications, especially given the high cost of both types of errors. Providing simple average point estimates of metrics should not be considered an adequate standard. The understanding of variation that business processes generate, and that the resulting financial statements reflect, should be common knowledge. It is important to note that illusions of control or chaos are not known to some individuals and are much more sinister behaviors than overconfidence. Fifth, large effect sizes, like those shown for segment six in the SPC POS-PLS analysis, would be needed to get the attention of the business community. If an effect size of 0.25 is educationally significant, it would not be large enough to create development impetus. Minimum effect sizes need to be those used for clinical trials of 0.50 to 0.60 to attract interest. Both the SPC and true-false POS-PLS models contained certain segments with considerable effect sizes exceeding those for clinical trials.

Limitations

This study had a few limitations. The presence of two endogenous variables allowed for a verification of generalizability within the field of accounting and finance; however, beyond this group, there is no claim to a broader reflection of the knowledge within other business functional groups or to a population at large. This research used a Qualtrics panel, but Amazon Mechanical Turks or perhaps other specialized areas of finance and accounting with specific statistical training might have generated different results. If extant, the additional identification of the attributes of the segmentation results from the POS-PLS would have augmented understanding. Likewise, perhaps some readers might criticize the design of the survey as either too complex or too simple to make a valid assessment. Further analysis of the screened responses, speeders, or other non-compliant respondents could have yielded supplementary or confirmatory information.

Future Research

Many future research possibilities lie ahead. Objective assessments, those allowing respondents to answer the presented graph or proffered true-false questions without being able to gauge a correct answer, would be preferred to typical Likert-scaled questionnaires. The experiments in the psychology of illusion of control require priming of individuals; however, no attempt to induce that condition occurred in this study. Therefore, similar assessments could and should be developed for future trials as a method to thwart socially-desirable responding. A simpler version of the 18-string assessment could be attempted by using a 12-string SPC chart. Shorter strings would reduce the non-random rule count by two which would give subjects fewer data points to evaluate. Likewise, an international cohort could be explored. Qualtrics offered panels in the United Kingdom and India as possible large sample studies of greater than 1,000 respondents.

Since the costs of both types of errors can be large, efforts to create objective measures that evaluate all individuals for both illusion of control, as advocated by Fenton-O'Creevy et al. (2003), and illusion of chaos, as advocated by Harris and Osman (2012), should be pursued. Surprisingly, this effort has not been seized on by the academic community.

Taking a reductive approach at an individual level, microphenomenology offers a path forward for both training subjects and, with the help of an interviewer, revealing the tacit knowledge that is currently unknown to them. While Klein (1993) and others, in their research of recognition primed decision making, left tacit knowledge as a partial black box, microphenomenology might pierce the mystery of that implicit intelligence for random sequence perception as it has with artists and medical patients.

The exploration of groups two, five, and six found among the respondents and cross-referenced between the two assessments should be explored further. Reflecting on the insights from these segments could provide a path to awareness of many psychological vulnerabilities and opportunities for personal improvement for all individuals. Moreover, the discovery of these segments using these SEM-PLS techniques that perform rigorous exploratory research within a single software package augments any scientist's toolbox.

Closing Thoughts

Two final thoughts concerning this dissertation project follow. This study was written in a straightforward manner so that business personnel might read, comprehend the results, and draw their own implications of the facts as they were explained. And secondly, the rigors of this study in both sampling and analyses should refute the tendency within the finance and accounting community to react to the results as, "How can this be so?" Some individuals whether inside or outside the profession, will likely criticize the lack of statistical knowledge found among respondents in the study as a one-off and perhaps attempt to dismiss the results.

References

Ackert, L. F., Church, B. K., & Qi, L. (2015). An Experimental Examination of Portfolio Choice. *Review of Finance, 20, July 2016,* 1427–1447. https://doi.org/10.1093/rof/rfv036.

Ackert, L. F. & Deaves, R. (2011). *Understanding Behavioral Finance.* Delhi, India: South-Western, a part of Cengage Learning.

Anderegg, W.R., Callaway, E.S., Boykoff, M.T., Yohe, G. & Root, T.L. (2014). Awareness of Both Type 1 and 2 Errors in Climate Science and Assessment. *Bull. American Meteorological Society, 95,* 1445–1451. https://doi.org/10.1175/BAMS-D-13-00115.1

Armstrong, R. L. (1987). The Midpoint on a 5-point Likert Scale. *Perpetual and Motor Skills, 64,* 359-362.

Astrachan, C. A., Patel, V. K., & Wanzenried, G. (2014). A Comparative Study of CE-SEM and PLS-SEM for Theory Development in Family Firm Research. *Journal of Family Business Strategy, 5,* 116-128.

Ayton, P. & Fischer, I. (2004). The Hot Hand Fallacy and the Gambler's Fallacy: Two faces of subjective randomness? *Memory & Cognition (2004) 32, 8,* 1369-1378. https://doi.org/10.3758/BF03206327

Ayton, P., Hunt, A. J., & Wright, G. (1989). Psychological conceptions of randomness. *Journal of Behavioral Decision Making, 2, 4,* 221-238. http://dx.doi.org/10.1002/bdm.3960020403

Bar-Hillel, M., & Wagenaar, W. (1991). The Perception of Randomness. *Advances in Applied Mathematics, 12,* 428-454.

Bennedsen, M., & Henry B. (2016). *Henokiens: The Families and Firms Who Made History.* Fountainebleau, France: Insead.

Bennett, B. (2019). Ludic Fallacy. Website: logicallyfallacious.com. Downloaded from https://www.logicallyfallacious.com/tools/lp/Bo/LogicalFallacies/121/Ludic-Fallacy

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

Biggs, S. F., & Wild, J. J. (1985). An Investigation of Auditor Judgment in Analytical Review. *The Accounting Review, 4,* 607-633.

Bishop, J., & Carlin, J. B. (1993). Bias, Prevalence and Kappa. *Journal of Clinical Epidemiology, 46, 5,* 423-429. DOI: 10.1016/0895-4356(93)90018-V

Blair, E. & Zinkhan, G. M. (2006). Nonresponse and Generalizability in Academic Research. *Journal of the Academy of Marketing Science, 34, 1,* 4-7.

Blais, A-R., Thompson, M. M., & Baranski, J. V. (2005). Individual Differences in Decision Processing and Confidence Judgments in Comparative Judgment Tasks: The Role of Cognitive Styles. *Personality and Individual Differences, 38, 7,* 1701-1713. https://doi.org/10.1016/j.paid.2004.11.004

Buhl, E. R., Goodson, P., & Neilands, T. B. (2007). Structural Equation Modeling: A Primer for Health Behavior Researchers. *American Journal of Health Behavior, 31, 1,* 74-85.

Burger, J. M., & Cooper, H. M. (1979). The Desirability of Control. *Motivation and Emotion, 3, 4,* 381-393.

Burton, R. (2005). Multiple-Choice and True-False Tests: Myths and Misapprehensions. *Assessment & Evaluation in Higher Education, 30,1,* 65-72.

Bye, J., Fraedrich, K., Kirk, E., Schubert, S., & Zhu, Hiuhua. (2011). Random walk lengths of about 30 years in global climate. *Geophysical Research Letters, 38, 5,* 1-3. https://doi.org/10.1029/2010GL046333

Byrt, T., Bishop, J. & Carlin, J.B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology, 46, 5,* 423-429. https://doi.org/10.1016/0895-4356(93)90018-V

Cabeza, R., Rao, S. M., Wagner, A. D., Mayer, A. R., & Schacter, D. L. (2001). Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study of veridical and illusory recognition memory. *Proceedings of the National Academy of Sciences, 98, 88,* 4805-4810.

Christensen, L. B., Johnson, R. B., & Turner, L. A. (2007). *Research Methods, Design, Analysis*. Upper Saddle River, NJ: Pearson Education Inc.

Cicchetti, D. V., & Feinstein, A. R. (1990). High Agreement but Low Kappa: II. Resolving the Paradoxes. *Journal of Clinical Epidemiology, 43, 6,* 511-558.

Christensen, L. B., Johnson, R. B., & Turner, L. A. (2014). *Research Methods, Design, and Analysis.* Boston, MA: Pearson Education.

Chyung, S.Y., Roberts, K., Swanson, & I. Hankinson, A. (2017). Evidence-Based Survey Design: The Use of a Midpoint on the Likert Scale. *Performance Improvement, 56, 10,* 15-23.

Clemmer, J. (1992). *Firing on all Cylinders. The Service/Quality System for High-Powered Corporate Performance*. Toronto, Canada: Macmillan Canada.

Clow, K. A., & Esses, V. M. (2005). The Development of Group Stereotypes from Descriptions of Group Members: An Individual Difference Approach. *Group Processes & Intergroup Relations, 8, 4,* 429-445. https://doi.org/10.1177/1368430205056469

Costa, D. F., De Melo Carvalho, F., & de Melo Moreira, B. C. (2019). Behavioral Economics and Behavioral Finance: A Bibliometric Analysis of the Scientific Fields. *Journal of Economic Surveys, 33, 1,* 3-24.

Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* Thousand Oaks, CA: Sage Publications.

Davidson, A., & Laroche, M. M. (2016). Connecting the dots: how personal need for structure produces false consumer pattern perceptions. *Marketing Letters, 27, 3,* 337-350. https://doi.org/10.1007/s11002-014-9332-z

Davidson, R. A. (1986). Source of Funding and Outcome of Clinical Trials. *Journal of General Internal Medicine, 1,* 155-158.

DeCarlo, L. T. (2005). A Model of Rater Behavior in Essay Grading Based on Signal Detection Theory. *Journal of Educational Measurement, 42, 1*, 53-76. doi:10.1111/j.0022-0655.2005.00004.x

DeCarlo, L. T., Kim, Y. & Johnson, M. S. (2011). A Hierarchical Rater Model for Constructed Responses, with a Signal Detection Rater Model. *Journal of Educational Measurement, 48, 3,* 333-356. doi:10.1111/j.1745-3984.2011.00143.x

Delsole, T., & Tippett, M. K. (2016). Forecast Comparison Based on Random Walks. *American Meteorological Society, February 2016*.

Dembroski, T. M., MacDougall, J. M., & Musante, L. (1984). Desirability of Control versus Locus of Control: Relationship to Paralinguistics in the Type A interview. *Health Psychology, 3, 1,* 15-26. http://dx.doi.org/10.1037/0278-6133.3.1.15

Deming, W. E. (1982). *Out of the Crisis*. Cambridge, MA: Massachusetts Institute of Technology.

Deshmukh, A., Karim, K. E., Siegal, P. H. (1998). An Analysis of Efficiency and Effectiveness of Auditing to Detect Management Fraud: A Signal Detection Theory Approach. *International Journal of Auditing, 2,* 127-138.

De Vany, A. (2004). *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry.* New York, NY: Routledge.

Dieguez, S., Wagner-Egger, P., & Gauvrit, N. (2015). Nothing Happens by Accident, or Does It? A Low Prior for Randomness Does Not Explain Belief in Conspiracy Theories. *Psychological Science, 2015, 26, 11, Supplemental Material.* DS1-DS15.

Dimock, M. (2019). Defining generations: Where Millennials end and Generation Z begins. *Pew Research Center*. Retrieved from https://pewrsr.ch/2szqtJz

Dopuch, N., Holthausen, R. W., & Leftwich, R. W. (1987). Predicting Qualification with Financial and Market Variables. *The Accounting Review, 62, 3,* 431-454.

Dunn, K. A., Tan, C. E. L., & Venuti, E. K. (2012). Audit Firm Characteristics and Type II Errors in the Going-Concern Opinion. *Asia-Pacific Journal of Accounting & Economics, 9, 1,* 39-69.

Ebert, J. P. & Wenger, D. M. (2011). Mistaking Randomness for Free Will. *Consciousness and Cognition 20, 3,* 965-971.

Ellis, R. S. (2000). Entropy as a Measure of Randomness; A Fundamental concept in the Mathematical Sciences. Fifteenth Annual Mathematic Competition Awards Ceremony, April 27, 2000.

Ericsson A. (2003). Valid and Non-Reactive Verbalization of Thoughts During Performance of Tasks Towards a Solution to the Central Problems of Introspection as a Source of Scientific Data. *Journal of Consciousness Studies, 10,* 1-18.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review, 104, 2,* 301-318.

Fast, N. J., Gruenfeld, D. H., Sivanathan, N., & Galinsky, A. D. (2008). Illusory Control: A Generative Force Behind Power's Far-Reaching Effects, Research Paper No. 2009, Stanford Graduate School of Business.

Feinstein, A. R., & Cicchetti, D. V. (1990). High Agreement but Low Kappa: I. Resolving the Paradoxes. *Journal of Clinical Epidemiology, 43, 6,* 511-558.

Fenton-O'Creevy, M., Nicholson, N., Soane, E., & Willman, P. (2003). Trading on Illusions: Unrealistic perceptions of Control and Trading Performance. *Journal of Occupational and Organizational Psychology, 76,* 53-68.

Filippin, A., & Crosetto, P. (2015). Click'n'Roll: No Evidence of Illusion of Control. Discussion Paper No, 9030.

Fitza, M. A. (2017). How Much do CEOs Really Matter? Reaffirming that the CEO Effect is Mostly Due to Chance. *Strategic Management Journal, 38,* 802–811. DOI: 10.1002/smj.2597

Fitza, M. A. (2014). The Use of Variance Decomposition in the Investigation of CEO Effects: How Large Must the CEO Effect Be to Rule Out Chance? *Strategic Management Journal, 35,* 1839-1852. DOI: 10.1002/smj.2192

Forstmeier, W., Wagenmakers, E., & Parker, T. H. (2017). Detecting and Avoiding Likely False-Positive Findings – a Practical Guide. *Biological Reviews, 92,* 1941-1968. https://doi.org/10.1111/brv.12315

Freiman, J. A., Chalmers, T. C., Smith, H. Jr., & Kuebler, R. R. (1978). The Importance of Beta, the Type II Error and Sample Size in the Design and Interpretation of the Randomized Control Trial. *The New England Journal of Medicine, 299,* 690-694.

Fritz, M. (1921). Guessing in a True-False Test. *Journal of Educational Psychology, 18, 8,* 558-561.

Funabashi, H. (2009). *Timeless Ventures: 32 Japanese Companies that Imbibed 8 Principles of Longevity.* Deli, India: Tata McGras-Hill Publishing Company Limited.

Garland, R. (1991). The Mid-Point on a Rating Scale: Is it Desirable? *Marketing Bulletin, 2,* 66-70, Research Note 3.

Garoupa, N. (1997). A Note on Private Enforcement and Type I Error. *International Review of Law and Economics, 17,* 423-429.

Gebhardt, W. A., & Brosschot, J. F. (2002). Desirability of control: psychometric properties and relationships with locus of control, personality, coping, and mental and somatic complaints in three Dutch samples. *European Journal of Personality, 16,* 423-438. https://doi.org/10.1002/per.463

Geiger, M. A., & Rama, D. V. (2006). Audit Firm Size and Going-Concern Reporting Accuracy. *Accounting Horizons, 20, 1,* 1-17.

Gerlach, M., Farb, B., Revelle, W., & Amaral, L. A. N. (2018). A Robust Data-driven Approach Identifies Four Personality Types Across Four Large Data Sets. *Nature of Human Behavior, 2,* 735–742. https://doi.org/10.1038/s41562-018-0419-z

Giakoumakis, E., Papaconstantinou, G., Skordalakis, E. (1987). Rule-Based Systems and Pattern Recognition, *Pattern Recognition Letters, 5, 4,* 267-272.

Gigerenzer, G. (1991). From Tools to Theories: A Heuristic of Discover in Cognitive Psychology. *Psychological Review, 98, 2,* 265-267.

Gigerenzer, G. (2000). *Adaptive Thinking: Rationality in the Real World.* Oxford University Press. New York, New York.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest, 8, 2,* 53–96. https://doi.org/10.1111/j.1539-6053.2008.00033.x

Gladwell. M. (2005). *Blink.* New York, NY: Back Bay Books, Little, Brown, and Company, Hachette Book Group.

Golder, P. N., & Tellis, G. J. (1993). Pioneer Advantage: Marketing Logic or Marketing Legend? *Journal of Marketing Research, 30, 2,* 158-170.

Griffiths, D., Bunder, M., Gulati, C., & Onizawa, T. (2010). The Probability of an Out of Control Signal from Nelson's Supplementary Zig-Zag Test. *Centre for Statistical and Survey Methodology, University of Wollongong, Working Paper 11-10,* 1-9. http://ro.uow.edu.au/cssmwp/61

Grossmann, I., & Kross, E. (2014). Exploring Solomon's Paradox: Self-Distancing Eliminated the Self-Other Asymmetry in Wise Reasoning About Close Relationships in Younger and Older Adults. *Psychological Science, 25, 8,* 1-10. https://doi.org/10.1177/0956797614535400

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM).* Thousand Oaks, CA: Sage Publications.

Hair, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2018). *Advanced Issues in Partial Least Squares Structural Equation Modeling*. Thousand Oaks, CA: Sage Publishing, Inc.

Harris A. J. L., & Osman, M. (2012). The Illusion of Control: A Bayesian Perspective. *Synthesis, 189,* 29-38.

Henderson, G. R., Mead, G. E., van Dijke, M. L., Ramsay, R., McDowall, M. A., & Dennis, M. (2008). Use of Statistical Process Control Charts in Stroke Medicine to Determine if Clinical Evidence and Changes in Service Delivery Were Associate with Improvements in the Quality of Care. *Quality Safety Health Care, 2008, 17,* 301-306.

Heuer, J., Merkel, C., & Weber, M. (2016). Fooled by Randomness: Investor Perception of Fund Manager Skill. *Review of Finance, 2016,* 605-635.

Hussein, A. (2009). The Use of Triangulation in Social Sciences Research: Can Qualitative and Quantitative be Combined? *Journal of Comparative Social Work, 1,* 1-12.

Ioannidis, J. P. A. (2005) Why Most Published Research Findings Are False. *PLOS Medicine, 2, 8.* e124. https://doi.org/10.1371/journal.pmed.0020124

Johnson, E. N. (1994). Auditor Memory for Audit Evidence: Effects of Group Assistance, Time Delay, and Memory Task. *Auditing: A Journal of Practice and Theory 13, 1,* 1-15.

Kahneman, D. & Klein, G. (2009). Conditions for Intuitive Expertise. A Failure to Disagree. *American Psychologist, 64, 6,* 515-526.

Kahneman, D., & Riepe, M. W. (1998). Aspects of Investor Psychology. *Journal of Portfolio Management, Summer 1998,* 52-65.

Kahneman, D. & Tversky A. (1972). Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology, 3,* 430-454.

Kasama.com. (2019). The 8 Wastes of Lean. Downloaded from https://www.kasama.us/8-wastes-of-lean/

Keys, D. E., & Reding, K. F. (1992). Statistical Process Control: What Management Accountants Need to Know. *Management Accounting, 73, 7,* 26-30.

Khan, M. M. S., Ghayas, M. M., & Kashif, S. (2019). Why Firms Fail to Sustain? Evidence from Dow Jones Index. *South Asian Journal of Management Science, 13, 2,* 116-136.

Kim, S., Eberhart, R., & Armanios, D. (2016). What Makes a Winner? Quantifying Luck in Sustained Performance. Working Paper, November 2016. DOI: 10.13140/RG.2.2.19222.14401

Kimhi, S., & Zysberg, L. (2009). How People Understand Their World: Perceived Randomness of Rare Life Events. *The Journal of Psychology, 2009, 143, 5,* 521-532.

Kinney, W. R. (1975). A Decision-Theory Approach to the Sampling in Auditing. *Journal of Accounting Research, 13, 1,* 117-132.

Kinney, W. R., & Salamon, G. L. (1982). Regression Analysis in Auditing: A Comparison of Alternative Investigation Rules. *Journal of Accounting Research. 20, 2,* 350-366.

Klarner, P., Sarstedt, M., Hoeck, M., & Ringle, C.M. (2013) Disentangling the Effects of Team Competences, Team Adaptability, and Client Communication on the Performance of Management Consulting Teams, *Long Range Planning, 46, 3,* 258-286. https://doi.org/10.1016/j.lrp.2013.03.001.

Klein, G. (1993). A Recognition Primed Decision (RPD) Model of Rapid Decision Making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds). *Decision Making in Action.* 138-147. AblexEditors.

Kordes, U. (2012). Thinking of Experience, Experiencing Thinking. *Interdisciplinary Description of Complex Systems, 10, 3,* 223-234.

Kriz, J. (1993). Pattern Formation in Complex Cognitive Systems. *Interdisciplinary Approaches, to nonlinear Complex Systems, 62,* 161-175.

Krueger, W. C. F. (1933). Distributions of Scores Based on Correct Guessing for True-false Tests of Various Lengths. *Journal of Educational Psychology 24, 3,* 185-188. DOI: 10.1037/h0072930

Kvålseth, T. O. (2015). Measurement of Interobserver Disagreement: Correction of Cohen's Kappa for Negative Values. *Journal of Probability and Statistics, 2015, 751803,* 1-8. http://dx.doi.org/10.1155/2015/751803

Kyriazos, T. A. (2018). Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psychology, 9,* 2207-2230.

Langer, E. J. (1975). The Illusion of Control. *Journal of Personality and Social Psychology, 32, 2,* 311-328.

Leanmanufacturingtools.com. (nd). The Waste of Talent or Creativity. Downloaded from https://leanmanufacturingtools.org/133/waste-of-talent-or-creativity-causes-symptoms-examples-solutions/

Leansmarts.com. (nd). A Summary of the Eight Wastes of Lean. Downloaded from https://leansmarts.com/lean-101/8-wastes/

Lobato, E., Mendoza, J., Sims, V. & Chin, M. (2014). Examining the Relationship Between Conspiracy Theories, Paranormal Belief, and Pseudoscience Acceptance Among a University Population. *Applied Cognitive Psychology, 28,* 617-625.

Lopes, L. L., & Oden, G. C. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 3,* 392-400. http://dx.doi.org/10.1037/0278-7393.13.3.392.

Lorenz, E.N. (1963). Deterministic Non-periodic Flow. *Journal of Atmospheric Science, 20,* 130-141.

Mandelbrot, B. (1977). *The Fractal Geometry of Nature.* New York, NY: W.H. Freeman and Company.

Mandelbrot, B., & Hudson, R. L. (2004). *The (Mis)Behavior of Markets. A Fractal View of Risk, Ruin, & Reward.* New York, NY: Basic Books.

Mandelbrot, B., & Taleb, N. N. (2010). Mild vs. Wild Randomness: Focusing on Those Risks that Matter. In F. Diebold, N. Doherty, & R. Herring. (Eds.), *The Known, the Unknown and the Unknowable in Financial Institutions,* 47-58. Princeton, NJ: Princeton University Press.

McDonald, F. E. J. (2009). *Understanding Randomness via the Perception and Prediction of Binary Sequences.* (Doctoral dissertation, School of Psychology, University of New South Wales. Retrieved from https://trove.nla.gov.au/work/37717374?q&versionId=252598102

McHugh M. L. (2012). Interrater Reliability: The Kappa Statistic. *Biochemical Medicine, 22,* 3, 276-82.

McSweeney, B. (2006). Net Present Value: the Illusion of Certainty. *Strategic Change, 15,* 47-51.

Mlodinow, L. (2008). *The Drunkard's Walk: How Randomness Rules Our Lives.* New York, NY: Vintage Books Random House.

Mohammed, M. A., Cheng, K. K., Rouse, A., & Marshall, T. (2001). Bristol, Shipman, and Clinical Governance: Shewhart's Forgotten Lessons. *The Lancet, 357*, 463-467.

Morrison D. G. (1978). A Probability Model for Forced Binary Choices. *The American Statistician, 32, 1,* 23-25. http://www.jstor.org/stable/2683470

Mauboussin, M. J., Callahan, D., & Majd, D. (2017). Corporate Longevity: Index Turnover and Corporate Performance. Website: www.credit-suisse.com. Downloaded from https://researchdoc.creditsuisse.com/docView?language=ENG&format=PDF&sourceid=em&document_id

Nelson, L. S. (1984). The Shewhart Control Chart—Tests for Special Causes. *Journal of Quality Technology, 16, 4,* 237-239. DOI: 10.1080/00224065.1984.11978921

Nickerson, R. S. (2002). The Production and Perception of Randomness. *Psychological Review, 109, 2,* 330–357.

Noskievičová, D. (2013). Complex Control Chart Interpretation. *International Journal of Engineering Business Management, 5, 13,* 1-7. https://doi.org/10.5772/56441

Novovic, Z., Kovac, A., Duric, B., & Biro, M. (2012). Positive and Negative Affect in Illusion of Control. *Psihologija, 45, 4,* 395-407.

O'Brien, P. C., & Fleming, T. R. (1979). A Multiple Testing Procedure for Clinical Trials. *Biometrics, 35,* 549-556.

Ohtsuka, K., & Ohtsuka, T. (2010). Vietnamese Australian Gamblers' Views on Luck and Winning: Universal Versus Culture-specific Schemas. *Asian Journal of Gambling Issues and Public Health, 1, 34,* 34-46. https://doi.org/10.1186/BF03342117

Olivola, C. Y., & Oppenheimer, D. M. (2008). Randomness in Retrospect: Exploring the Interactions between memory and Randomness Cognition. *Psychonomic Bulletin & Review, 15, 5,* 991-996.

Overgaard, M., Gallagher, S., & Ramsoy, T. Z. (2008), An Integration of First-Person Methodologies in Cognitive Science. *Journal of Consciousness Studies: Controversies in Science and the Humanities, 15, 5,* 100-120.

Parnell, J. A., & Dent, E. B. (2009). The Role of Luck in the Strategy-Performance Relationship. *Management Decision, 47, 6,* 1000-1021. https://doi.org/10.1108/00251740910966703

Peterson, G. T. (1998). The Influence of C. I. Lewis on Shewhart and Deming. *Excelsis Business Advisers,* October 1998. Retrieved from http://www.scrummaster.dk/lib/AgileLeanLibrary/People/CILewis/C-I-Lewis-Shewhart-and-Deming-by-G-T-Peterson.pdf

Petitmengin, C. (2006). Describing One's Subjective Experience in the Second person: An Interview Method of the Science of Consciousness, *Phenomenology and Cognitive Science, 5,* 229–269. https://doi.org/10.1007/s11097-006-9022-2.

Petitmengin, C. (2019). Studying the Experience of Meditation Through Micro-Phenomenology, *Current Opinion in Psychology, 28,* 54-59. https://doi.org/10.1016/j.copsyc.2018.10.009

Pi, Y., Liao, W., Lui, N. & Lu, J. (2008). Theory of Cognitive Pattern Recognition. In P. Yin (Ed.), *Pattern Recognition Techniques, Technology and Applications,* 433-463. Vienna, Austria: InTech Journals.

Poundstone, W. (2014). *Rock Breaks Scissors: A Practical Guide to Outguessing & Outwitting Almost Anybody.* New York, NY: Little, Brown and Company, Hachette Book Group.

Presson, P. K., & Benassi, V. A. (1996). Illusion of Control: A Meta-Analytic Review. *Journal of Social Behavior & Personality, 11, 3,* 493-510.

Quigley, T. J., & Graffin, S. D. (2017). Reaffirming the CEO Effect is Significant and Much Larger than Chance: A Comment on Fitza (2014). *Strategic Management Journal, 38,* 793–801. DOI: 10.1002/smj.2503

Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review, 77, 6,* 481-495. http://dx.doi.org/10.1037/h0029964

Rietzschel, E. F., De Dreu, C. K. W., & Nijstad, B. A. (2007). Personal Need for Structure and Creative Performance: The Moderating Influence of Fear of Invalidity. *Personality and Social Psychology Bulletin, 33, 6,* 855–866. https://doi.org/10.1177/0146167207301017

Rizzolli, M. (2016). Adjudication Type-I and Type-II Errors. Working paper 15. *Center for Relationship Banking and Economics, December 2016,* 1-14.

Rodin, J., & Langer. E. J. (1977) Long-Term Effects of a Control-Relevant Intervention with the Institutionalized Aged. *Journal of Personality and Social Psychology, 35, 12,* 897-902.

Ross, K. G., Klein, G. A., Thunholm, P., Schmitt. J. F., & Baxter, H. C. (2004). The Recognition-Primed Decision Model. *Military Review, July-August 2004,* 6-10.

Salas. E., Rosen, M. A., & DiazGranados, D. (2010). Expertise-Based Intuition and Decision Making in Organizations. *Journal of Management* Online First. DOI: 10.1177/0149206309350084

Sale, C. (1929). *The Specialist.* St. Louis, MO: Specialist Publishing Company, Wainwright Building.

Sanderson, Y. B. (2009). Effective generation of subjectively random binary sequences. *Advances in Applied Mathematics, 43, 1,* 1-11.

Sheehan, K. B., & Pittman. M. (2016). *Amazon's Mechanical Turk for Academics: The HIT Handbook for Social Science Research*. Irvine, CA: Melvin and Leigh Publishers.

Scheibehenne, B., Wilke, A., & Todd, P. M. (2011). Expectations of clumpy resources influence predictions of sequential events. *Evolution and Human Behavior, 32,* 326–333.

Schnotz, W., Wagner, I. (2018). Construction and Elaboration of Mental Models Through Strategic Conjoint Processing of Text and Pictures. *Journal of Educational Psychology, January 22, 2018.* Advance online publication. http://dx.doi.org/10.1037/edu0000246

Selden, P. H. (1996). *Sales Process Engineering: A Personal Workshop*. Milwaukee, WI: ASQC Quality Press.

Shewhart, W. E. (2015). *Economic Control of Quality of Manufactured Product*. Mansfield Center, CT: Martino Publishing. (Originally printed in 1931 by D. Van Nostrad Company, Inc.).

Shumla.com (2007). The 8th Type of Muda. Downloaded from https://www.shmula.com/the-8th-type-of-muda/309/

Silver, N. (2012). *The Signal and the Noise: Why so Many Predictions Fail – but Some Don't.* New York, NY: Penguin Books.

Skhmot, N. (2017). The 8 Wastes of Lean. Downloaded from https://theleanway.net/The-8-Wastes-of-Lean

Stimmler, Mary. (2013). Essays on Financial Risk Taking and Embedded Heuristics. UC Berkeley: Business Administration, Ph.D. Program. Retrieved from: http://www.escholarship.org/uc/item/01z2b0hj

Sorensen, J. E. (1969). Bayesian Analysis in Auditing. *American Accounting Association. 44, 3,* 555-561.

Soyer, E., & Hogarth, R. M. (2012). The Illusion of Predictability: How regression Statistics Mislead Experts. *International Journal of Forecasting, 28,* 695-711.

SPC Excel. (n.d.) Control Chart Rules and Interpretation. Retrieved from https://www.spcforexcel.com/knowledge/control-chart-basics/control-chart-rules-interpretation

SPC Excel. (n.d.) Over-controlling a Process: The Funnel Experiment. Retrieved from https://www.spcforexcel.com/knowledge/variation/overcontrolling-process-funnel-experiment

Steenkamp;, J. B. E. M., De Jong, M. G., & Baumgartner, H. (2010). Socially Desirable Response Tendencies in Survey Research. *Journal of Marketing Research, 47, 2,* 199-214.

Stöckl, T., Huber, J. M., Kirchler, M., & Lindner, F. (2005). Hot hand and Gambler's fallacy in Teams: Evidence from Investment experiments. *Journal of Economic Behavior & Organization, 117,* 327-339.

Stojanov, A., & Halberstadt, H. (2020). Does Lack of Control Lead to conspiracy Belief? A Meta-Analysis. *European Journal of Social Psychology, 2020, 00,* 1– 14. https://doi.org/10.1002/ejsp.2690

Sullivan, G.M., & Feinn, R. (2012) Using Effect Size – or Why the *P-value* I not Enough, *Journal of Graduate Medical Education, September 2012,* 279-282.

Svecova, V., & Pavlovicova, G. (2016). Screening the Personal Need for the Structure and solving word problems with fractions. *SpringerPlus, 5,* 652-561. doi:10.1186/s40064-016-2285-3

Taleb, N. N. (2004). *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets.* New York, NY: Random House.

Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable.* New York, NY: Random House.

Taleb, N. N. (2008). Errors, Robustness, and The Fourth Quadrant. Retrieved from http://static.stevereads.com/papers_to_read/errors_robustness_and_the_fourth_quadrant.pdf

Taleb, N. N. (2014). *AntiFragile: Things that Gain from Disorder.* New York, NY: Random House.

Taleb, N. N., & Goldstein, D. G. (2011). The Problem is Beyond Psychology: The Real World is More Random than Regression Analyses (October 10, 2011). *International Journal of Forecasting, Forthcoming,* 1-2. Downloaded from https://ssrn.com/abstract=1941792

Taylor, P. J., & Small, B. (2002). Asking Applicants What They Would Do Versus What They Did Do: A Meta-Analytic Comparison of Situational and Past Behaviour Employment Interview Questions. *Journal of Occupational and Organizational Psychology, 75,* 277-294.

Tellis, G. J. (2013). *Unrelenting Innovation. How to Build a Culture of Market Dominance.* San Francisco, CA: John Wiley and Sons.

Thompson, M. M., Naccarato, M. E., Parker, K. C. H., & Moskowitz, G. (2001). The Personal Need for Structure (PNS) and Personal Fear of Invalidity (PFI) scales: Historical Perspectives, Present Applications and Future Directions. *In G. Moskowitz (Ed.), Cognitive Social Psychology: The Princeton Symposium on the Legacy and Future of Social Cognition,* 19-39. Mahwah, NJ: Erlbaum.

Tyszka T., Zielondka, P., Dacey, R., & Sawicki, P. (2008) Perception of randomness and predicting uncertain events. *Thinking & Reasoning, 14,1,* 83-110, DOI: 10.1080/13546780701677669

Tversky A., & Kahneman, D. (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science, 185, 457,* 1124-1131.

Uecker, W. C., & Kinney, Jr., W. R. (1977). Judgmental Evaluation of Sample Results: A Study of the Type and Severity of Error Made by Practicing CPAs. *Accounting, Organizations, and Society. 2,3,* 269-275.

VanderWeft, P. A., & Mahon, J. F. (1997). Meta-Analysis of the Impact of Research Methods on Findins of Frist-Mover Advantage. *Management Science, 43, 11,* 1510-1519.

Viera, A. J., & Garrett, J. M. (2005). Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine, 37, 5,* 360-363.

von Eye, A., & von Eye, M. (2005). Can One Use Cohen's Kappa to Examine Disagreement? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1, 4,* 129-142. http://dx.doi.org/10.1027/1614-2241.1.4.129

Wagenaar, W. A., & Keren, G.B. (1988). Chance and luck are not the same. *Journal of Behavioral Decision Making, 1, 2,* 65-75. DOI: 10.1002/bdm.3960010202

Waggner-Egger, P., & Bangerter, A. (2007). La Vérité est Ailleurs : Corrélats de L'Adhésion aux Théories du Complot. *Presse Universitaires de Grenoble, 4,20,* 31-61.

Walton, M. (1990). *Deming Management at Work: Six successful Companies that use the Quality Principles of the world-famous W. Edwards Deming*. New York, NY: Penguin Books, The Putnam Publishing Company.

Wang, C., & Hill, D. J. (2007). Deterministic Learning and Rapid Dynamical Pattern Recognition. *IEEE Transactions on Neural Networks, 18, 3,* 617-630.

Warrens, M. J. (2014). New Interpretations of Cohen's Kappa. *Journal of Mathematics, 2014, 203907,* 1-9. http://dx.doi.org/10.1155/2014/203907

Wessling, K. S., Huber, J., & Netzer, O. (2017). MTurk Character Misrepresentation: Assessment and Solutions. *Journal of Consumer Research, 44,* 211-230.

Wheeler, D. J., & Chambers. D. C. (1992). *Understanding Statistical Process Control.* (2^{nd} Ed.). Knoxville, TN: SPC Press.

Whitson, J., Kim, J, Wang, C. S., Menon, T., & Webster, B. D. (2018). Regulatory Focus and Conspiratorial Perceptions: The Importance of Personal Control. *Personality and Social Psychology Bulletin,* 1-13.

Williams, J. J., & Griffiths, T. L. (2013). Why Are People Bad at Detecting Randomness? A Statistical Argument. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 39, 5,* 1473–1490.

Whitson, J., & Galinsky, A. (2008). Lacking Control Increases Illusory Pattern Perception. *Science, 322, 115*. DOI: 10.1126/science.1159845.

Xu, J., & Harvey, N. (2004). Carry on winning: The gamblers' fallacy creates hot hand effects in online gambling, *Cognition, 131, 2*, 173-180.

Zbilut, J. P. (2004). *Unstable Singularities and Randomness: Their Importance in the Complexity of Physical, Biological, and Social Sciences*. Elsevier. Amsterdam, The Netherlands.

Zhao, J., Hahn, U., & Osherson, D.O. (2014). Perception and Identification of Random Events. *Journal of Experimental Psychology ,40, 4,* 1358-1371.

Zysberg, Z., & Kimhi, S. (2011). Winning of Losing a Bet and the Perception of Randomness. *Journal of Gambling Studies*. DOI 10.1007/s10899-011-9289-2

Appendix A

Table A1.
Pilot Statistical Process Control Chart (SPC) Question List

Survey	**Answer**	**Error**	**Rule Violation; 0 = no Rule Violation**
SPC1	Random	Type1	0
SPC2	Non-Random	Type2	Test 6: 4 out of 5 points more than 1 standard deviation (StDev) from center line (CL) (same side)
SPC3	Random	Type1	0
SPC4	Non-Random	Type2	Test 1: 1 point more than 3 standard deviation from center line
SPC5	Random	Type1	0
SPC6	Random	Type1	0
SPC7	Random	Type1	0
SPC8	Non-Random	Type2	Test 1: 1 point more than 3 standard deviation from center line
SPC9	Random	Type1	0
SPC10	Non-Random	Type2	Test 8: 8 points in a row more than 1 standard deviation from CL (either side)
SPC11	Random	Type1	0
SPC12	Non-Random	Type2	Test 2: 7 points in a row on same side of center line
SPC13	Non-Random	Type2	Test 3: 7 points in a row all increasing or all decreasing
SPC14	Non-Random	Type2	Test 7: 14 points in a row within 1 standard deviation from center line (either side)
SPC15	Random	Type1	0
SPC16	Non-Random	Type2	Test 4: 14 points in a row alternating up and down
SPC17	Random	Type1	0
SPC18	Non-Random	Type2	Test 2: 7 points in a row on same side of center line
SPC19	Random	Type1	0
SPC20	Non-Random	Type2	Test 5: 2 out of 3 points more than 2 standard deviation from center line (same side)

Note. Statistical process control charts answer keys which includes question number, correct answer, the type of error if incorrect, and the Nelson's rule violation if any. The order of the list was randomized.

Table A2.
Pilot True-False (T-F) Question List

Survey	Answer	Error	Actual Question
TF1	FALSE	Type1	The Picture of Dorian Grey, chapter 4, 32th word is "with".
TF2	TRUE	Type2	The Picture of Dorian Grey, chapter 2, 14th word is "with".
TF3	TRUE	Type2	Gunman's Reckoning, chapter 26, 20th word is "the".
TF4	FALSE	Type1	Gunman's Reckoning, chapter 2, 11th word is "time".
TF5	TRUE	Type2	The Little Price, Karen Woods' translation, chapter 9, 6th word is "escape".
TF6	FALSE	Type1	Gunman's Reckoning, chapter 26, 20th word is "distance".
TF7	TRUE	Type2	Lamentations 3:5, Chouraqui Bible; 4th word is "moi".
TF8	TRUE	Type2	The Picture of Dorian Grey, chapter 4, 32th word is "room".
TF9	FALSE	Type1	Riders of the Purple Sage, chapter 3, 11th word is "men".
TF10	TRUE	Type2	Qohelet 1:5, Chouraqui Bible; 8th word is "son".
TF11	FALSE	Type1	Riders of the Purple Sage, chapter 5, 21th word is "the".
TF12	TRUE	Type2	Gunman's Reckoning, chapter 2, 11th word is "from".
TF13	FALSE	Type1	The Little Price, Karen Woods' translation, chapter 4, 11th word is "importance".
TF14	FALSE	Type1	Qohelet 1:5, Chouraqui Bible; 8th word is "lieu".
TF15	FALSE	Type1	The Picture of Dorian Grey, chapter 2, 14th word is "his".
TF16	FALSE	Type1	Lamentations 3:5, Chouraqui Bible; 4th word is "et".
TF17	TRUE	Type2	The Little Price, Karen Woods' translation, chapter 4, 11th word is "this".
TF18	FALSE	Type1	The Little Price, Karen Woods' translation, chapter 9, 6th word is "his".
TF19	TRUE	Type2	Riders of the Purple Sage, chapter 5, 21th word is "on".
TF20	TRUE	Type2	Riders of the Purple Sage, chapter 3, 11th word is "some".

Note. The true-false answer keys include question number, correct answer, the type of error if incorrect, and the actual question. The order of the list was randomized.

Table A3.
Pilot Desirability of Control (DC) Question List

Number	Question	Coded
1	I prefer a job where I have a lot of control over what I do and when I do it.	
2	I enjoy political participation because I want to have as much of a say in running government as possible.	
3	I try to avoid situations where someone else tells me what to do.	
4	I would prefer to be a leader than a follower.	
5	I enjoy being able to influence the actions of others.	
6	I am careful to check everything on an automobile before I leave for a long trip.	
7	Others usually know what is best for me.	
8	I enjoy making my own decisions.	
9	I enjoy having control over my own destiny.	
10	I would rather someone else take over the leadership role when I'm involved in a group project.	**R**
11	I consider myself to be generally more capable of handling situations than others are.	
12	I'd rather run my own business and make my own mistakes than listen to someone else's orders.	
13	I like to get a good idea of what a job is all about before I begin.	
14	When I see a problem, I prefer to do something about it rather than sit by and let it continue.	
15	When it comes to orders, I would rather give them than receive them.	
16	I wish I could push many of life's daily decisions off on someone else.	**R**
17	When driving, I try to avoid putting myself in a situation where I could be hurt by another person's mistake.	
18	I prefer to avoid situations where someone else has to tell me what it is I should be doing.	
19	There are many situations in which I would prefer only one choice rather than having to make a decision.	**R**
20	I like to wait and see if someone else is going to solve a problem so that I don't have to be bothered with it.	**R**

Note. The desirability of control question list which includes question number, question, and coding type (either straight or reverse).

Table A4.
Pilot Personal Need for Structure (PNS) Question List

Number	Question	Coded
1	It upsets me to go into a situation without knowing what I can expect from it.	
2	I'm not bothered by things that upset my daily routine.	**R**
3	I enjoy having a clear and structured mode of life.	
4	I like a place for everything and everything in its place.	
5	I like being spontaneous.	**R**
6	I find that a well-ordered life with regular hours makes my life tedious.	**R**
7	I don't like situations that are uncertain.	
8	I hate to change my plans at the last minute.	
9	I hate to be with people that are unpredictable.	
10	I find that a consistent routine enables me to enjoy life more.	
11	I enjoy the exhilaration of being put in unpredictable situations.	**R**
12	I become uncomfortable when the rules in a situation are not clear.	

Note. The personal need for structure question list which includes question number, question, and coding type (either straight or reverse).

Table A5.
Pilot Bipolar Characteristics Question List

Number	Anchor 1	Anchor 2
C1	Look What I Avoided	Look What I Did
C2	Small losses with a small probability of a large gain	Small gains with a small probability of a large loss
C3	Medium losses with a medium probability of large gains	Medium gains with a medium probability of large losses
C4	Trial-and-error	Tried-and-True
C5	Follower strategy	First to market, first mover strategy

Note. The bipolar characteristics question list includes the question number and both anchors.

Table A6.
Pilot Distressed Characteristics Question List

Number	Question
C6	How much has participating in a financially distressed company affected your business perspective? (financial distress includes major reorganization, credit workout, rescue, bankruptcy, liquidation)
C7	How much have particular life stresses affected your business perspective?

Note. The distressed characteristics question list includes the question number and question.

Table A7.
Pilot Demographic Question List

Number	**Question**
C8	Generation: Baby boomer, Gen X, Gen Y, Silent, Greatest
C8	Highest level of education completed: high school, college, graduate school
C9	Gender: male, female

Note. The demographics question list includes the question number and question.

Appendix B

Table B1.
Final Statistical Process Control Chart (SPC) Question List

Survey	Answer	Error	Rule Violation; 0 = no Rule Violation
SPC1	Random	Type1	0
SPC2	Non-Random	Type2	Test 6: 4 out of 5 points more than 1 standard deviation (StDev) from center line (CL) (same side)
SPC3	Random	Type1	0
SPC4	Non-Random	Type2	Test 1: 1 point more than 3 standard deviation from center line
SPC5	Random	Type1	0
SPC6	Random	Type1	0
SPC7	Random	Type1	0
SPC8	Non-Random	Type2	Test 1: 1 point more than 3 standard deviation from center line
SPC9	Random	Type1	0
SPC10	Non-Random	Type2	Test 8: 8 points in a row more than 1 standard deviation from CL (either side)
SPC11	Random	Type1	0
SPC12	Non-Random	Type2	Test 2: 7 points in a row on same side of center line
SPC13	Non-Random	Type2	Test 3: 7 points in a row all increasing or all decreasing
SPC14	Non-Random	Type2	Test 7: 14 points in a row within 1 standard deviation from center line (either side)
SPC15	Random	Type1	0
SPC16	Non-Random	Type2	Test 4: 14 points in a row alternating up and down
SPC17	Random	Type1	0
SPC18	Non-Random	Type2	Test 2: 7 points in a row on same side of center line
SPC19	Random	Type1	0
SPC20	Non-Random	Type2	Test 5: 2 out of 3 points more than 2 standard deviation from center line (same side)

Note. Statistical process control charts answer keys which includes question number, correct answer, the type of error if incorrect, and the Nelson's rule violation if any. The order of the list was randomized. No changes from pilot.

Table B2.
Final True-False (T-F) Question List

Survey	Answer	Error	Actual Question
TF1	FALSE	Type1	The Picture of Dorian Grey, chapter 4, 32th word is "with".
TF2	TRUE	Type2	The Picture of Dorian Grey, chapter 2, 14th word is "with".

TF3	TRUE	Type2	Gunman's Reckoning, chapter 26, 20th word is "the".
TF4	FALSE	Type1	Gunman's Reckoning, chapter 2, 11th word is "time".
TF5	TRUE	Type2	The Little Price, Karen Woods' translation, chapter 9, 6th word is "escape".
TF6	FALSE	Type1	Gunman's Reckoning, chapter 26, 20th word is "distance".
TF7	TRUE	Type2	Lamentations 3:5, Chouraqui Bible; 4th word is "moi".
TF8	TRUE	Type2	The Picture of Dorian Grey, chapter 4, 32th word is "room".
TF9	FALSE	Type1	Riders of the Purple Sage, chapter 3, 11th word is "men".
TF10	TRUE	Type2	Qohelet 1:5, Chouraqui Bible; 8th word is "son".
TF11	FALSE	Type1	Riders of the Purple Sage, chapter 5, 21th word is "the".
TF12	TRUE	Type2	Gunman's Reckoning, chapter 2, 11th word is "from".
TF13	FALSE	Type1	The Little Price, Karen Woods' translation, chapter 4, 11th word is "importance".
TF14	FALSE	Type1	Qohelet 1:5, Chouraqui Bible; 8th word is "lieu".
TF15	FALSE	Type1	The Picture of Dorian Grey, chapter 2, 14th word is "his".
TF16	FALSE	Type1	Lamentations 3:5, Chouraqui Bible; 4th word is "et".
TF17	TRUE	Type2	The Little Price, Karen Woods' translation, chapter 4, 11th word is "this".
TF18	FALSE	Type1	The Little Price, Karen Woods' translation, chapter 9, 6th word is "his".
TF19	TRUE	Type2	Riders of the Purple Sage, chapter 5, 21th word is "on".
TF20	TRUE	Type2	Riders of the Purple Sage, chapter 3, 11th word is "some".

Note. The true-false answer keys include question number, correct answer, the type of error if incorrect, and the actual question. The order of the list was randomized. No changes from pilot.

Table B3.
Final Desirability of Control (DC) Question List

Number	**Question**	**Coded**
1	I prefer a job where I have a lot of control over what I do and when I do it.	
2	I enjoy political participation because I want to have as much of a say in running government as possible.	
3	I try to avoid situations where someone else tells me what to do.	
4	I would prefer to be a leader than a follower.	
5	I enjoy being able to influence the actions of others.	
6	I am careful to check everything on an automobile before I leave for a long trip.	
7	Others usually know what is best for me.	
8	I enjoy making my own decisions.	
9	I enjoy having control over my own destiny.	
10	I would rather someone else take over the leadership role when I'm involved in a group project.	**R**

11	I consider myself to be generally more capable of handling situations than others are.	
12	I'd rather run my own business and make my own mistakes than listen to someone else's orders.	
13	I like to get a good idea of what a job is all about before I begin.	
14	When I see a problem, I prefer to do something about it rather than sit by and let it continue.	
15	When it comes to orders, I would rather give them than receive them.	
16	I wish I could push many of life's daily decisions off on someone else.	**R**
17	When driving, I try to avoid putting myself in a situation where I could be hurt by another person's mistake.	
18	I prefer to avoid situations where someone else has to tell me what it is I should be doing.	
19	There are many situations in which I would prefer only one choice rather than having to make a decision.	**R**
20	I like to wait and see if someone else is going to solve a problem so that I don't have to be bothered with it.	**R**

Note. The desirability of control question list which includes question number, question, and coding type (either straight or reverse). No changes from pilot.

Table B4.

Final Personal Need for Structure (PNS) Question List

Number	Question	Coded
1	It upsets me to go into a situation without knowing what I can expect from it.	
2	I'm not bothered by things that upset my daily routine.	**R**
3	I enjoy having a clear and structured mode of life.	
4	I like a place for everything and everything in its place.	
5	I like being spontaneous.	**R**
6	I find that a well-ordered life with regular hours makes my life tedious.	**R**
7	I don't like situations that are uncertain.	
8	I hate to change my plans at the last minute.	
9	I hate to be with people that are unpredictable.	
10	I find that a consistent routine enables me to enjoy life more.	
11	I enjoy the exhilaration of being put in unpredictable situations.	**R**
12	I become uncomfortable when the rules in a situation are not clear.	

Note. The personal need for structure question list which includes question number, question, and coding type (either straight or reverse). No changes from the pilot questionnaire.

Table B5.
Final Bipolar Characteristics Question List

Number	Anchor 1	Anchor 2
C1	Look What I Avoided	Look What I Did
C2	Small losses with a probability of a large gain	Small gains with a probability of a large loss
C3	Medium losses with a probability of large gains	Medium gains with a probability of large losses
C4	Tried-and-True	Trial-and-Error
C5	Follower strategy	First to market, first mover strategy

Note. The bipolar characteristics question list includes the question number and both anchors. Changes from pilot were to remove the probability qualifiers of small and medium and reverse the positions of tried-and-true and trial-and-error.

Table B6.
Final Distressed Characteristics Question List

Number	Question
C6	How much has participating in a financially distressed company affected your business perspective? (financial distress includes major reorganization, credit workout, rescue, bankruptcy, liquidation)
C7	How much have particular life stresses affected your business perspective?

Note. The distressed characteristics question list includes the question number and question. No changes from pilot.

Table B7.
Pilot Demographic Question List

Number	Question
C8	Birth year
C8	Highest level of education completed: high school, college, graduate school
C9	Gender: male, female

Note. The demographics question list includes the question number and question. Birth year added and fixed generational categories removed.

Table B8.

Final Bipolar Characteristics Question List after Third Trial

Number	Anchor 1	Anchor 2
C1	Look What I Avoided	Look What I Did
C2	Consistent small losses with a probability of a large gain	Consistent small gains with a probability of a large loss
C3	Consistent medium losses with a probability of large gains	Consistent medium gains with a probability of large losses
C4	Tried-and-True	Trial-and-Error
C5	Follower strategy	First to market, first mover strategy

Note. The bipolar characteristics question list includes the question number and both anchors. Changes from third trial were to add the qualifier of consistent to the losses to gains or gains to losses bipolar questions.

Appendix C

Table C1.
Outer loadings from Initial Model

Indicator	Char Type1	Char Type2	DC	PFI	SPCType2 Error Rate	SPCType1 Error Rate
C1	0.295					
C1R		0.286				
C2	0.871					
C2R		0.933				
C3	0.531					
C3R		0.560				
C4	0.610					
C4R		0.380				
C5	0.490					
C5R		0.438				
DC1			-0.335			
DC10			0.614			
DC11			-0.448			
DC12			-0.570			
DC13			-0.231			
DC14			-0.143			
DC15			-0.534			
DC16			0.757			
DC17			-0.156			
DC18			-0.542			
DC19			0.696			
DC2			-0.597			
DC20			0.721			
DC3			-0.604			
DC4			-0.425			
DC5			-0.475			
DC6			-0.362			
DC7			0.695			
DC8			-0.118			
DC9			-0.160			
PFI1				-0.468		
PFI10				-0.502		
PFI11				0.572		
PFI12				0.608		
PFI13				-0.633		
PFI14				0.720		
PFI2				-0.376		
PFI3				0.650		

PFI4	0.701		
PFI5	0.726		
PFI6	0.726		
PFI7	0.580		
PFI8	0.459		
	-		
PFI9	0.668		
SPCType1			1.000
SPCType2		1.000	

Note. Outer loadings greater than 0.7 exhibit convergent validity.

Table C2.

Outer Weights (Formative) for Initial Model

Relationship	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (|O/STDEV|)	*P-values*
C1 -> Char Type1	0.182	0.172	0.185	0.982	0.326
C1R -> Char Type2	0.184	0.182	0.158	1.158	0.247
C2 -> Char Type1	**0.742**	**0.692**	**0.172**	**4.320**	**0.000**
C2R -> Char Type2	**0.868**	**0.820**	**0.143**	**6.074**	**0.000**
C3 -> Char Type1	-0.038	-0.043	0.221	0.173	0.863
C3R -> Char Type2	0.025	0.022	0.196	0.129	0.897
C4 -> Char Type1	0.291	0.274	0.192	1.520	0.129
C4R -> Char Type2	-0.002	0.003	0.166	0.010	0.992
C5 -> Char Type1	0.291	0.278	0.174	1.670	0.095
C5R -> Char Type2	0.284	0.271	0.158	1.792	0.074

Note. Outer weights test for significance and *p-values*. Significant values are shown in bold.

Table C3.

Reliability Metrics for the Initial Model

Measure	Cronbach's Alpha	Composite Reliability	Average Variance Extracted (AVE)
DC	0.795	0.247	0.253
PFI	0.454	0.521	0.371
SPCType2 Error Rate	1.000	1.000	1.000
SPCType1 Error Rate	1.000	1.000	1.000

Note. Cronbach's alpha should be a minimum value of 0.7.

Table C4.
Reliability Metrics for the Final Model

Measure	Cronbach's Alpha	Composite Reliability	Average Variance Extracted (AVE)
DC	0.852	0.894	0.629
PFI	0.861	0.896	0.589
SPCType2 Error Rate	1.000	1.000	1.000
SPCType1 Error Rate	1.000	1.000	1.000

Note. Cronbach's alpha should be a minimum value of 0.7.

Table C5.
Fornell-Larker Criteria for Discriminant Validity for the Initial Model

Measure	Char Type1	Char Type2	DC	PFI	SPCType2 Error Rate	SPCType1 Error Rate
Char Type1						
Char Type2	-0.961					
DC	-0.349	0.383	**0.503**			
PFI	0.333	-0.346	-0.781	**0.609**		
SPCType2 Error	0.176	-0.183	-0.224	0.219	1.000	
SPCType1 Error	-0.161	0.155	0.216	-0.213	-0.690	1.000

Note. DC and PFI have highest row values. Bold values indicate acceptable discriminant validity.

Table C6.
Fornell-Larker Criteria for Discriminant Validity for the Final Model

Measure	Char Type1	Char Type2	DC	PFI	SPCType2 Error Rate	SPCType1 Error Rate
Char Type1 Error						
Char Type2 Error	-0.934					
DC	-0.378	0.412	0.793			
PFI	0.365	-0.371	-0.758	0.768		
SPCType2 Error	0.164	-0.171	-0.208	0.191	1.000	
SPCType1 Error	-0.151	0.141	0.196	-0.174	-0.690	1.000

Note. DC and PFI have highest row values.

Table C7.

Hetero-Trait Mono-Trait (HTMT) Criteria for Discriminant Validity for the Initial Model

Measure	DC	PFI	SPCType2 Error Rate
PFI	0.772		
SPCType2 Error Rate	0.171	0.224	
SPCType1 Error Rate	0.166	0.216	0.690

Note. In a complete bootstrapping procedure, confidence intervals for single measures.

Table C8.
Hetero-Trait Mono-Trait (HTMT) Criteria for Discriminant Validity for the Final Model

Measure	DC	PFI	SPCType2 Error Rate
PFI	0.874		
SPCType2 Error Rate	0.224	0.201	
SPCType1 Error Rate	0.211	0.183	0.690

Note. In a complete bootstrapping procedure, confidence intervals for single measures.

Table C9.
Structural Model Assessment with relationships and P-values.

Relationship	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (|O/STDEV|)	*P-Values*
Char Type1 -> SPCType1	**-0.085**	**-0.093**	**0.032**	**2.628**	**0.009**
Char Type2 -> SPCType2	**-0.099**	**-0.101**	**0.033**	**2.995**	**0.003**
DC -> SPCType2 Error	**-0.118**	**-0.120**	**0.050**	**2.387**	**0.017**
DC -> SPCType1 Error	0.131	0.130	0.048	2.711	0.007
PFI -> SPCType2 Error	0.065	0.068	0.051	1.289	0.198
PFI -> SPCType1 Error	-0.043	-0.046	0.049	0.888	0.375

Note. Significant values are shown in bold.

Table C10.
Path Coefficients and P-values for Kappa Group Six.

Measures	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (|O/STDEV|)	*P-Values*
Char Type1 -> SPCType1	**-0.107**	**-0.119**	**0.039**	**2.731**	**0.007**
Char Type2 -> SPCType2	**-0.111**	**-0.115**	**0.044**	**2.530**	**0.012**
DC -> SPCType2 Error	-0.044	-0.045	0.061	0.718	0.473
DC -> SPCType1 Error	0.043	0.046	0.058	0.748	0.455
PFI -> SPCType2 Error	0.087	0.096	0.056	1.567	0.118
PFI -> SPCType1 Error	-0.054	-0.058	0.057	0.951	0.342

Note. Group six was composed of respondents without under or over-alternation bias in true-false assessment. Significant values are shown in bold.

Table C11.
Path Coefficients and P-values for Kappa Group 1.

Measures	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (|O/STDEV|)	*P-Values*
Char Type1 -> SPCType1	-0.211	-0.152	0.217	0.975	0.330
Char Type2 -> SPCType2	-0.166	-0.168	0.132	1.260	0.208
DC -> SPCType2 Error	-0.172	-0.217	0.208	0.829	0.407
DC -> SPCType1 Error	0.203	0.251	0.199	1.020	0.308
PFI -> SPCType2 Error	0.106	0.094	0.211	0.504	0.615
PFI -> SPCType1 Error	-0.047	-0.025	0.189	0.247	0.805

Note. Group one straight-lined all True on true-false assessment.

Table C12.

Path Coefficients and P-values for Kappa Group 2.

Measures	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (|O/STDEV|)	*P-Values*
Char Type1 -> SPCType1	-0.076	-0.113	0.085	0.893	0.373
Char Type2 -> SPCType2	-0.065	-0.077	0.091	0.717	0.474
DC -> SPCType2	-0.033	-0.037	0.129	0.259	0.796
DC -> SPCType1	0.113	0.113	0.134	0.844	0.399
PFI -> SPCType2	0.149	0.159	0.122	1.215	0.225
PFI -> SPCType1	-0.073	-0.086	0.133	0.545	0.586

Note. Group two exhibited True tendencies with under-alternation on true-false assessment.

Table C13.

Path Coefficients and P-values for Kappa Group 3

Measures	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (|O/STDEV|)	*P-values*
Char Type1 -> SPCType1	**-0.267**	**-0.290**	**0.095**	**2.815**	**0.005**
Char Type2 -> SPCType2	**-0.294**	**-0.301**	**0.112**	**2.631**	**0.009**
DC -> SPCType2	-0.084	-0.092	0.184	0.456	0.649
DC -> SPCType1	0.121	0.107	0.218	0.556	0.578
PFI -> SPCType2	0.124	0.115	0.182	0.683	0.495
PFI -> SPCType1	-0.069	-0.046	0.185	0.375	0.707

Note. Group three exhibited False tendencies with under-alternation on true-false assessment. Significant values are shown in bold.

Table C14.

Path Coefficients and P-values for Kappa Group 4

Measures	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (|O/STDEV|)	*P-values*
Char Type1 -> SPCType1	0.172	0.208	0.141	1.222	0.222
Char Type2 -> SPCType2	0.183	0.202	0.136	1.349	0.178
DC -> SPCType2	0.018	0.060	0.179	0.103	0.918
DC -> SPCType1	0.006	-0.042	0.160	0.036	0.972
PFI -> SPCType2	-0.233	-0.213	0.224	1.040	0.299
PFI -> SPCType1	0.230	0.198	0.215	1.072	0.284

Note. Group three straight-lined False with under-alternation on true-false assessment.

Table C15.

Path Coefficients and P-values for True Tendencies on the True-False Assessment

	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (|O/STDEV|)	*P-values*
Char Type1 -> SPCType1	-0.093	-0.104	0.044	2.106	0.036
Char Type2 -> SPCType2	-0.070	-0.080	0.047	1.500	0.134
DC -> SPCType2	-0.074	-0.070	0.075	0.975	0.330
DC -> SPCType1	0.129	0.128	0.073	1.777	0.076
PFI -> SPCType2	**0.140**	**0.148**	**0.068**	**2.062**	**0.040**
PFI -> SPCType1	-0.099	-0.106	0.067	1.474	0.141

Note. Significant values are shown in bold.

Table C16.

Fit Indexes for One- to Eight-Segment Solution (FIMIX-PLS)

Fit Index/Segment Number	1	2	3	4	5	6	7	8
AIC (Akaike's Information Criterion)	5,868	5,647	5,506	5,448	5,382	5,335	5,324	**5,305**
AIC3 (Modified AIC with Factor 3)	5,876	5,664	5,532	5,483	5,426	5,388	5,386	**5,376**
AIC4 (Modified AIC with Factor 4)	5,884	5,681	5,558	5,518	5,470	**5,441**	5,448	5,447
BIC (Bayesian Information Criteria)	5,907	5,731	5,635	5,621	5,600	**5,597**	5,631	5,657
CAIC (Consistent AIC)	5,915	5,748	5,661	5,656	**5,644**	5,650	5,693	5,728
HQ (Hannan Quinn Criterion)	5,883	5,679	5,555	5,514	5,465	**5,434**	5,441	5,439
MDL5 (Minimum Description Length with Factor 5)	**6,130**	6,204	6,359	6,595	6,825	7,072	7,357	7,633
LnL (LogLikelihood)	(2,926)	(2,807)	(2,727)	(2,689)	(2,647)	(2,614)	(2,600)	(2,582)
EN (Entropy Statistic (Normed))		0.496	0.658	0.514	0.707	0.643	0.652	0.544
NFI (Non-Fuzzy Index)		0.533	0.643	0.461	0.649	0.555	0.547	0.406
NEC (Normalized Entropy Criterion)		529	359	510	308	375	365	479

Note. Row minimum is optimal and shown in bold. Preferred combinations: AIC_3/CAIC, AIC_3/BIC, AIC_4/BIC. EN should be larger than 0.50. This solution used AIC_4/BIC.

Table C17.

Relative Segment Sizes (percentage of sample n=1,050) by FIMIS-PLS

Number of Segments/Size	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6	Segment 7	Segment 8
2	0.712	0.288						
3	0.703	0.217	0.080					
4	0.535	0.224	0.164	0.077				
5	0.664	0.124	0.093	0.063	0.056			
6	0.549	0.153	0.144	0.056	0.050	0.048		
7	0.522	0.160	0.105	0.071	0.058	0.057	0.027	
8	0.280	0.280	0.120	0.100	0.088	0.073	0.044	0.015

Note. Segment 6: groups after segment one range between 50 and 160 respondents.

Table C18.

R^2 for the Original Sample and Six Segments (FIMIX-PLS)

Relationship/Segment	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6	Original Sample R-Squares
SPCType2 Error Rate	0.137	0.458	0.862	0.944	0.965	0.989	0.054
SPCType1 Error Rate	0.097	0.463	0.803	0.869	0.960	0.993	0.046

Note. Segments three to six exhibit high R^2.

Table C19.

Path Coefficients for Original Sample and Six POS Segments (POS-PLS)

Relationship	Original	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6
Char Type1 -> Type1	**-0.085**	**-0.132**	**0.086**	**1.065**	**-1.002**	**0.155**	0.016
Char Type2 -> Type 2	**-0.099**	**-0.147**	0.034	**0.997**	**-0.993**	**0.162**	**0.058**
DC -> Type1	**0.131**	**0.217**	**0.202**	-0.015	-0.142	**-0.529**	**1.513**
DC -> Type2	**-0.118**	**-0.202**	**-0.107**	0.118	**0.218**	**0.503**	**-1.508**
PFI -> Type1	-0.043	**0.118**	**-0.833**	-0.313	-0.079	**0.454**	**0.716**
PFI -> Type2	0.065	-0.084	**0.909**	0.281	0.170	**-0.484**	**-0.686**
Group Size	1,050	749	80	38	61	54	68

Note. Segment sizes are also listed. Significant values are shown in bold.

Table C20.

R^2 Original, by POS Segment, and Weighted (POS-PLS)

R^2	Original	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6	Weighted
SPCType1 Error	0.046	0.049	0.946	0.894	0.962	0.948	0.983	0.308
SPCType2 Error	0.054	0.060	0.965	0.908	0.947	0.961	0.986	0.318
Group Size	1,050	749	80	38	61	54	68	1,050

Note. Segment sizes are also listed.

Table C21.

Path Relationships Original and POS Segments

Relationships	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (|O/STDEV|)	*P-values*
Char Type1 -> SPCType1	0.025	0.019	0.023	1.071	0.285
Char Type2 -> SPCType2	**0.058**	**0.055**	**0.020**	**2.937**	**0.003**
DC -> SPCType1 Error	**1.514**	**1.512**	**0.082**	**18.480**	**0.000**
DC -> SPCType2 Error	**-1.508**	**-1.506**	**0.079**	**19.122**	**0.000**
PFI -> SPCType1 Error	**0.714**	**0.707**	**0.072**	**9.949**	**0.000**
PFI -> SPCType2 Error	**-0.686**	**-0.676**	**0.072**	**9.516**	**0.000**

Note. Compare this table to Appendix Table C9 of original model. Significant values are shown in bold.

Table C22.
Fit Indexes for One- to Nine-Segment Solution (FIMIX-PLS) for True-False Model

Fit Index/Segment Number	1	2	3	4	5	6	7	8	9
AIC	5,749	5,516	5,362	5,273	5,218	5,200	5,152	5,133	**5,116**
AIC3	5,757	5,533	5,388	5,308	5,262	5,253	5,214	5,204	**5,196**
AIC4	5,765	5,550	5,414	5,343	5,306	5,306	5,276	**5,275**	5,276
BIC	5,788	5,601	5,491	5,446	**5,436**	5,463	5,459	5,484	5,513
CAIC	5,796	5,618	5,517	5,481	**5,480**	5,516	5,521	5,555	5,593
HQ	5,764	5,548	5,411	5,338	5,301	5,300	5,268	**5,266**	5,266
MDL5	**6,011**	6,074	6,214	6,420	6,661	6,938	7,184	7,460	7,739
LnL	(2,866)	(2,741)	(2,655)	(2,601)	(2,565)	(2,547)	(2,514)	(2,495)	(2,478)
EN	0.000	0.411	0.427	0.474	0.468	0.509	0.489	0.579	0.574
NFI	0.000	0.450	0.443	0.439	0.413	0.423	0.389	0.455	0.436
NEC	0	619	602	553	559	515	537	443	447

Note. Row minimum is optimal and shown in bold. Preferred combinations: AIC_3/CAIC, AIC_3/BIC, AIC_4/BIC. EN should be larger than 0.50. Segment size issues prohibited any segments larger than six; therefore, five segments were selected.

Table C23.
R^2 Original, by POS Segment, and Weighted (POS-PLS) for True-False Model

R^2	Original	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Weighted
SPCType1 Error	0.106	0.510	0.251	0.871	0.720	0.776	0.887
SPCType2 Error	0.099	0.439	0.150	0.822	0.704	0.710	0.893
Group Size	1,050	566	148	113	131	92	1,050

Note. Segments three to five exhibit high R^2.

Table C24.
Path Coefficients for Original Sample and Five POS Segments (POS-PLS) for True-False Model

Relationship	Original	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
Char Type1 -> Type1	**0.078**	**0.271**	-0.093	**-0.854**	**0.799**	0.026
Char Type2 -> Type 2	0.041	-0.066	0.054	**-0.944**	**0.761**	**0.216**
DC -> Type1	**0.136**	**0.196**	0.030	-0.098	-0.224	**0.993**
DC -> Type2	**-0.177**	**-0.282**	-0.053	-0.064	0.085	**-0.889**
PFI -> Type1	**0.218**	**0.225**	**0.943**	0.099	-0.067	**-0.132**
PFI -> Type2	**-0.192**	**-0.181**	**-0.868**	**-0.226**	0.191	**0.122**
Group Size	1,050	566	148	113	131	92

Note. Segment sizes are also listed. Significant values are shown in bold.

Table C25.
Comparison of SPC and True-False Models Assignment of Respondents

Segments	TF1	TF2	TF3	TF4	TF5	Total
SPCGroup1	453	89	65	84	58	749
SPCGroup2	32	23	9	6	10	80
SPCGroup3	14	3	9	7	5	38
SPCGroup4	16	14	9	19	3	61
SPCGroup5	23	8	9	9	5	54
SPCGroup6	28	11	12	6	11	68
Total	566	148	113	131	92	1,050

Note. The SPC group adds horizontally while the true-false groups add vertically.