Students' Meanings for Stochastic Process

While Developing a Conception of Distribution

by

Neil J. Hatfield

A Dissertation Presented in Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy

Approved March 2019 by the
Graduate Supervisory Committee:

Patrick Thompson, Chair
Marilyn Carlson
James Middleton
Richard Lehrer
Mark Reiser

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

The concept of distribution is one of the core ideas of probability theory and inferential statistics, if not *the* core idea. Many introductory statistics textbooks pay lip service to stochastic/random processes but how do students think about these processes? This study sought to explore what understandings of stochastic process students develop as they work through materials intended to support them in constructing the long-run behavior meaning for distribution.

I collected data in three phases. First, I conducted a set of task-based clinical interviews that allowed me to build initial models for the students' meanings for randomness and probability. Second, I worked with Bonnie in an exploratory teaching setting through three sets of activities to see what meanings she would develop for randomness and stochastic process. The final phase consisted of me working with Danielle as she worked through the same activities as Bonnie but this time in teaching experiment setting where I used a series of interventions to test out how Danielle was thinking about stochastic processes.

My analysis shows that students can be aware that the word "random" lives in two worlds, thereby having conflicting meanings. Bonnie's meaning for randomness evolved over the course of the study from an unproductive meaning centered on the emotions of the characters in the context to a meaning that randomness is the lack of a pattern. Bonnie's lack of pattern meaning for randomness subsequently underpinned her image of stochastic/processes, leading her to engage in pattern-hunting behavior every time she needed to classify a process as stochastic or not. Danielle's image of a stochastic process was grounded in whether she saw the repetition as being reproducible (process can be

repeated, and outcomes are identical to prior time through the process) or replicable (process can be repeated but the outcomes aren't in the same order as before). Danielle employed a strategy of carrying out several trials of the process, resetting the applet, and then carrying out the process again, making replicability central to her thinking.

DEDICATION

This dissertation is dedicated to all of my Statistics students.

ACKNOWLEDGMENTS

Writing a dissertation is a labor of love that is not possible without a dynamic support system. For me, my support system includes my family, friends, and my mentors.

A thank you to my family who believed in me, asked the dreaded question of how things were going, and gave me the space I needed.

I must thank my fellow graduate students for all of their support and encouragement over the years. Without all of you, grad school would not have been nearly as fun nor nearly as enriching. In particular, there are a few colleagues who helped me shape my thinking: Dr. Sarah Burke for being my Statistics buddy and my sounding board when I began my journey of bringing quantitative reasoning to distributions and moments; Gabriel Tarr for being a Guinea pig for my applets; Morgan Sellars for allowing me to talk Statistics with her. To Ashley Duncan, I owe a big debt of gratitude: without you I don't know how I would have stayed sane. From our "date" nights to trivia and everything in between, you have been a key player in my support system.

I must also thank the wonderful individuals in the ASU SoMSS graduate and undergraduate offices: you rock! To rest of my ASU family as well as my current UNCO family, thank you for all of your support.

To my mentors, thank you for how you have each pushed me. Dr. Mary Shepherd: this dissertation is a result of you introducing me to a whole new world of mathematics education research; without you I wouldn't have gone on to do my PhD at ASU. Dr. Mark Reiser for helping to shape my understanding of Statistics and providing

me with opportunities to push the development of materials for Intro. Statistics. I owe thanks to Drs. Richard Lehrer, James Middleton, and Luis Saldanha for helping push my thinking in Statistics and Mathematics Education, reminding me that there are perspectives beyond radical constructivism. Dr. Marilyn Carlson, you have been a life-saver jumping on to my committee at the last minute, but more importantly, you've been a rock and cheerleader. Your support through the years, your willingness to chat with me (even late into the night) about wide ranging topics in higher education has helped shape me into who I am today.

Finally, my deepest gratitude goes to Dr. Patrick Thompson. You've pushed me to move beyond procedures and reason quantitatively. Starting back in that first semester of Calculus and encouraging Cameron and I dig into how the students were thinking about division, you've helped me grow as a researcher. Allowing me to be part of Project Aspire from the start was a phenomenal experience that had a direct impact on not only this dissertation but how I think about research. Your guidance both in terms of Statistics, Mathematics/Statistics Education, and being a faculty member are invaluable. While some people are intimidated by you, I know from first-hand experience that you are a true support: letting us get out our emotions, helping us to get the root of the current problem, and helping to put us back together again. Thank you for everything.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF EXCERPTS

Excerpt Page

Excerpt Page

Excerpt Page

Excerpt Page

Chapter 1: Introduction

This dissertation describes an investigation into how students understand the concept of a random variable's distribution. Serving as the bridge connecting the fields of Probability and Statistics, distribution is one of the most important ideas in Statistics. Without this concept, statistical inference would not exist in the form we know, if at all. Both parametric and non-parametric methods of inference rely on the idea of distribution. However, for such a critical concept, students' meanings for distribution have received little attention.

Statement of the Problem

Statistics courses have had increasing enrollments at both tertiary and high school levels. Multiple sets of standards have called for increased focus on statistics including the NCTM Standards (National Council of Teachers of Mathematics, 2000), the Guidelines for Assessment and Instruction in Statistics Education (GAISE) reports (Aliaga et al., 2005; Franklin et al., 2007), and the Common Core State Standards for Mathematics (CCSS-M) (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). While these standards documents have pushed statistical education forward, they treat "distribution" as a primitive/undefined term. The GAISE report for preK–12 stresses that "students understand the idea of a distribution" (Franklin et al., 2007, p. 23) but does not explain or exemplify what might constitute an understanding. Similarly, the GAISE college report notes that "students should understand the basic ideas of statistical inference including the concept of a sampling distribution" (Aliaga et al., 2005, p. 12). The CCSS-M states that students should "develop a probability distribution for a random variable" but stops short of suggesting how students should *think* about distributions or understand what it means to make one (see CCSS.Math.Content.HSS.MD.A.1–4).

The notion of distribution of a random variable entails the coordination of multiple ideas that are each complex in its own right. Students must bring together their images of randomness, random variable, random process, accumulation, and probability to build a coherent meaning for distribution. This coordination provides a basis upon which students can build and reason with sampling distributions for statistical inference. This is important for what Saldanha and Thompson (2014) refer to as the inner logic of statistical inference. Regardless of using replication, simulation (permutation, bootstrapping, or Monte Carlo), or asymptotic shortcut methods (traditional parametric or non-parametric), dealing with a distribution is inescapable for creating an inference. While much research exists on students' understandings of probability (e.g., Doerr, 2000; Konold, 1989; Saldanha & Liu, 2014), randomness (e.g., Falk & Konold, 1994; Kahneman & Tversky, 1982; Liu & Thompson, 2002; Metz, 1998), sample statistics such as the sample arithmetic mean (e.g., Clark, Kraut, Mathews, & Wimbish, 2007; Faradj, 2004; Mathews & Clark, 2007), little work has focused on how people think about distributions. Existing inquiry into students' meanings of distribution might be classified better as either 1) describing pictures of distributions (e.g., Arnold & Pfannkuch, 2012, 2014) or 2) using sampling distributions for inference (e.g., Jacob & Doerr, 2014; Lipson, 2003; Saldanha & Thompson, 2014). Given the centrality of the notation of distribution, investigating how students understand the distribution of random variable, beyond their ability to reason about graphs/images, as well as outside of inference is necessary. To help develop a statistically literate populace, we must understand how students understand distribution so that statistics education may start from what students understand in supporting their construction of more powerful understandings. This investigation aims to examine students' understandings of distribution of a random variable and difficulties they have in creating more powerful understandings. One of the most central aspects of the distribution concept is that of stochastic process. The meanings that students have for this idea will invariable shape their conception of distribution.

Research Questions

The notion of distribution is vast (see Chapter 4). To make this study more reasonable, I'm electing to use the distribution concept as a backdrop and focus in on stochastic processes. Thus, this study investigates students' meanings for stochastic process while they develop a conception of a random variable's distribution. The primary research question driving this study is:

- What meanings for stochastic process do students develop during an instructional sequence based upon a hypothetical learning progression for thinking of distribution describing the complete behavior of a stochastic process?

Secondary research questions that derive from the theoretical foundation and design of the study includes:

- What impact do students' meanings for randomness, random variable, and probability have on the development of their meaning for stochastic process during the instructional sequence?
- What images of accumulation do students develop during the instructional sequence?

Motivation for the Study

My personal experiences both as a statistician and as a statistics educator influenced this study. The concept of distribution in the courses I've taken was either unarticulated or was used as a label for a graph/table in the back of the textbook. As a student, I was content to have this notation be ill defined. Not having a productive meaning did not prevent me from succeeding in my courses. However, when I began teaching statistics, I ran into a problem. I struggled with how to help students understand the first two moments (expected value and variance) of a distribution in a way that was true to Thompson's theory of quantitative reasoning (Thompson, 1993, 2011). This is to say, that I wanted to help students understand that the first two moments measured an aspect of the distribution and have a sense of what a measurement value meant. After several discussions with a fellow statistics graduate student[1], I reached a point where I felt that I had something I could work with, but was still dissatisfied. I shared what I had done with Dr. Pat Thompson. His feedback helped me to see that my lingering dissatisfaction stemmed from the issue that the object whose attributes the moments

[1] I'm indebted to Sarah Burke for the countless hours she gave as a sounding board and the helpful feedback she gave me on my "wild" ideas for statistics education.

measured was not fully specified. The object was the distribution of a random variable. Upon this realization, the question became how to conceptualize the distribution of a random variable. Thus, began my investigation into how students understand the distribution of a random variable.

Chapter 2 provides a literature review for distribution along with ideas that it entails: ideas of randomness, random process and trial, random variable, and probability. Chapter 3 details the theoretical perspective that I used for the investigations. A conceptual analysis of the idea of distribution (of a random variable) along with progress variables and a hypothetical learning progression make up Chapter 4. I provide details and methods for my inquiry in the fifth chapter. Chapter 6 is the first of my result chapters, focusing on three students' conveyed meanings for randomness. I then describe one student's (Bonnie) meanings for stochastic process as Chapter 7. Chapter 8 looks at a second student (Danielle) and the meanings for stochastic process she developed. I conclude in Chapter 9 with a discussion of my results.

Chapter 2: Literature Review

Given that distribution entails the coordination of multiple ideas, I present the literature review in three sections. First, I will summarize the current literature on the concept of distribution. Then I'll present relevant literature on the ideas of randomness and random process. The final section will discuss individuals' understandings of probability.

Distribution

Existing research on students' understandings of distribution tends to fall into three focus areas: understanding distribution, describing [images of] distributions, and dealing with distributions in the service of another concept such as variation[2] or statistical inference. In an examination of 46 statistics texts, there are five categories for how the authors define the concept of distribution. Nine texts described distribution as some form of visual arrangement (e.g., a dot plot, histogram, or table) that shows the possible values along with the frequency of such a value. The most populous set of texts (18 texts) define distribution as being a pairing of the values of a random variable and either the frequency, relative frequency, or the probability for each value. While this category is similar to the first, there is no specific connection to data visualizations in this second set of texts. The third category (eight texts) focused on cumulative probability functions (distribution functions) and presented either an integral or the probability notation $P[X \leq x]$ without any discussion of what these symbols mean. Another eight texts make

[2] The concept of "variation" is nebulous and is often left undefined by authors, allowing each reader to impose whatever meaning they wish. When I use the word "variation" I mean noticed changes in the value of an attribute (see the discussion of random variable in Chapter 4). I will denote what I believe various authors to mean by "variation".

up the fourth category where the authors use the word "distribution" but do not say what they mean by this word. Finally, there are three texts that do not fit in any other categories; one equates distribution with a collection of data values and the other treats distribution as a synonym for frequency. The last text defines distribution in the following way: "The probability distribution of a random variable *x*, which we will denote by *F*, is any complete description of the probabilistic behavior of *x*" (Efron & Tibshirani, 1993, p. 22). The first two of these categories (i.e., visual arrangement and pairing of values) mimic what statistics education researchers seem to take as the meaning of distribution when they look at how students understand the distribution concept. The visual arrangement comes for viewing distribution as an arrangement while the pairing of values appears to stem for viewing distribution as a lens to examine an arrangement.

Distribution as arrangement. Leavy (2006) defined distribution as "the arrangement of values of a variable along a scale of measurement resulting in a representation of the observed or theoretical frequency of an event" (p. 90). Leavy investigated pre-service elementary teachers' understanding of distribution over a fifteen-week course. Leavy used the measures and representations that the teachers used to compare distributions as the teachers' understanding of distribution. At the onset of the course, the teachers did not appear to make use of distributional features (i.e., shape, center, variability [i.e. not all values being the same]/spread) when comparing data. While some teachers only made use of descriptive statistics such as the sample arithmetic mean, other teachers made use of graphical representations. Leavy noted that the graphics the teachers used at the beginning tended to illustrate various descriptive statistics rather than distributional features. At the end of the course, Leavy (2006) found that the teachers used graphical representations much more often and selected certain graphical representations (e.g. histograms, stem-and-leaf plots, parallel box plots) to highlight certain distributional features. Leavy took this shift as evidence that the teachers were now attending to global patterns in the data distributions. Leavy characterized the shift as the pre-service teachers changing their focus from summarizing to exploring.

Reading and Canada (2011) modified Leavy's statement by introducing probability in place of theoretical frequency. Reading and Canada (2011) go a step further and identify nine concepts upon which the notion of distribution depends: "center, variability [i.e., not the same], shape, density, skewness, relative frequency, probability, proportionality and causality" (p. 225). They argue that the first seven concepts are features of the distribution and that "center, variability (spread), and shape are commonly agreed (see, e.g., Bakker, 2004; Leavy, 2006; Pfannkuch & Reading, 2006; Shaughnessy, 2007; Garfield & Ben-Zvi, 2008) to be core concepts" (Reading & Canada, 2011, p. 225). These features of distribution appear to make up the concepts of empirical and theoretical distributions. The distinction between empirical and theoretical distributions lies in the use of density, skewness, and relative frequency for empirical and using probability for theoretical. The last two concepts (proportionality and causality) play a role in dealing with different sample sizes and making a connection between empirical and theoretical distributions.

Bakker and Gravemeijer (2004) expand on the notion of distribution as an arrangement with Figure 1. They argue that students typically start at the base of the structure (i.e., with individual data values) and move upwards. Experts combine both a downwards (i.e. starting with distribution) and upwards perspective. Bakker and Gravemeijer propose a three-stage framework for characterizing students' understandings of distribution. Their first stage is where students view a distribution as a set of data values. When students begin grouping data values together (e.g., creating bins for histograms or dot pots) and identifying modal clumps, Bakker and Gravemeijer say that the students are at the second stage. The final stage is when students begin to reason with modal clumps and majorities rather than individual values.

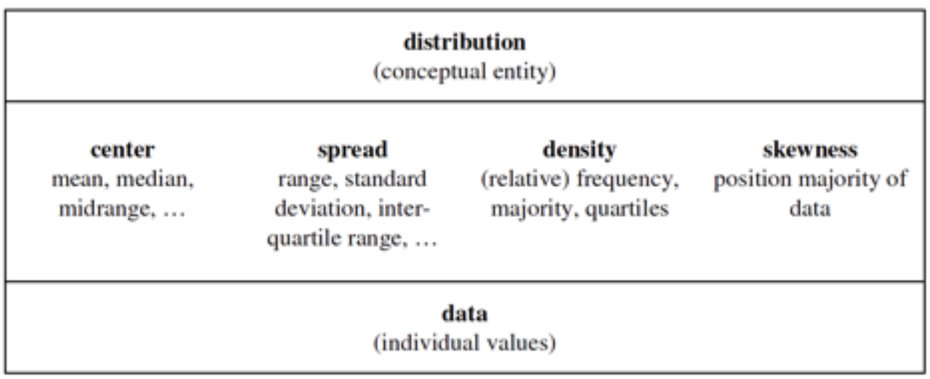


Figure 1. The structure of data and distribution (Bakker & Gravemeijer, 2004, p. 148).

In their closing remarks Garfield and Ben-Zvi (2004) define distribution as "a representation of quantitative data that can be examined and described in terms of shape, center, and spread, as well as unique features such as gaps clusters, outliers, and so on" (p. 400). This definition is consistent with Leavy's (2006) definition centering on an arrangement of values. Garfield and Ben-Zvi make explicit a focus on a graphical/imagistic representation that Leavy implies. The notion of distribution as an arrangement (or representation) of values appears to support instruction that places focus on attributes of distributions, in particular measures of center, spread, and shape. In turn these attributes of distribution become the very things that the distribution relies upon thus creating a circular relationship between the underlying object (the distribution) and the attributes of the object (center, spread, shape).

Viewing distribution as an arrangement or representation of values is not unique to statistics education research. Researchers focusing in judgment and decision-making and other fields also use this meaning of distribution. Goldstein and Rothschild (2014) used the Amazon Mechanical Turk labor market to survey 619 adults in an attempt to investigate the common adult's understanding of distribution. Goldstein and Rothschild focused how accurate an individual was in making forecasts about numeric information of a distribution. Participants watched a screen that displayed a sequence of one hundred numbered balls (numbered one to ten), one at a time, for 600 milliseconds each. Following this, the participants then used one of five methods to generate the distribution of numbers when they imagined resampling all 100 balls. The five methods Goldstein and Rothschild asked participants to use were a graphical method (drag and drop virtual balls into numbered bins), giving quantiles, estimating the arithmetic mean (formula given), estimating the "average" (no formula or definition of "average" given to participants), and to build a 90% confidence interval for the value of the first ball drawn when resampling. The last four methods Goldstein and Rothschild collapsed together under the title "standard method" to contrast with the graphical method. They found that at both the individual and aggregate levels, the graphical method resulted in more accurate responses. Goldstein and Rothschild (2014) suggest that these results support two hypotheses for what might be going on: 1) the individuals' mental representations of distribution are accurate (for the situation) but the standard methods cause the individuals to lose track of their representations, or 2) that the individuals' mental representations are inconsistent with the situation and the graphical method helps the individuals correct any heuristics/biases the individual used in the constructions of the mental representation. A second example is the work of Sheats and Pankratz (2002). The authors espouse two meanings for distribution. The first meaning is that distribution is the arrangement of data values into a graphical representation. The second meaning extends the first to equate "distribution" with "pattern". In both Goldstein and Rothschild (2014) and Sheats and Pankratz (2002), the authors take understanding distribution to mean identifying features of the arrangement or pattern of values.

Distribution as a lens. The second common meaning for distribution was put forth by Wild (2006). Written as the personal exploration of a statistician, Wild attempted to tackle what teachers and statistics education researchers want students to learn about the idea of distribution. He argued that at the core of the idea, distribution is "the pattern of variation in a variable" and that statisticians examine variation using distribution as a lens (Wild, 2006, p. 11).

Wild uses "pattern" to imply looking for and potentially removing regularities or trends in the data at hand. If a statistician removes all regularities in the data, then what is left is unexplained variation or noise (Wild, 2006). The regularities or trends (i.e., patterns) that a statistician removes, he attributes to what he understands about the way he generated the data and to particular sources/causes.

The issue of "variation" is much trickier; Wild (2006) wrote "In the beginning was variation. Variation is an observed reality detectable in all systems and entities. It is, in a word, omnipresent" (p. 10). Wild does not ever say what he means by "variation" in this article. Wild and Pfannkuch (1999) describe that "variation" can be real (part of the system a statistician studies) or induced (variation added via data collection, e.g. measuring, sampling, accidents). They convey that variation is the reason that "no two manufactured items are identical, no two organisms are identical or react in the identical ways…it is variation that makes the results of actions unpredictable, that makes questions of cause and effect difficult to resolve, that makes it hard to uncover mechanisms" (Wild & Pfannkuch, 1999, pp. 235–236). In essence, "variation" is what makes two (or more) things different and causes unpredictability.

While Wild never states what exactly he means by "lens", his use of the term conveys that a "lens" is a way of thinking about something we're examining. By drawing on Wild and Pfannkuch's (1999) elucidation of "variation" [i.e., lack of sameness] and Wild's (2006) meaning for "pattern", I believe that the best sense of what Wild means by his definition of distribution (i.e., "distribution is [a lens by which we examine] the pattern of variation in a variable") is that distribution is a way of thinking about data to see the regularity in what makes things different in a variable. This approach to defining the concept of distribution as a lens speaks directly to the use of distribution but falls short of saying anything about what makes up this concept. Treating distribution as a lens supports viewing statements such as “the random variable is Normally distributed zero, one” as a commitment to assumptions about the underlying random processes. However, for a student first starting out, he or she might not have the necessary schemes to support imbuing such a statement with a meaning consistent with a statistician does.

Magalhães (2014) reported on issues with how people understand the concept of distribution. While Magalhães (2014) does not give a preferred way of thinking about “distribution”, the author shared three problematic ways of thinking that pre-service mathematics teachers commonly had. The first two ways of thinking Magalhães described is where the students thought that random variables are completely unpredictable. Magalhães (2014) argued that students associate “random” with the idea of being uncontrolled or wild. This is not surprising, and I’ll return to this in the section on randomness. While this way of thinking is essentially about random variables, Wild and Pfannkuck’s (1999) notion of variation [i.e., the lack of sameness] includes viewing variation as the cause of unpredictability. If we take Wild’s (2006) notion of distribution, then thinking that random variables are unpredictable seems like a logical conclusion to viewing variables through a lens that is the “trend of the cause of unpredictability in a variable”. The second way of thinking that Magalhães (2014) described is that students believe that every value of the random variable has the same probability. Essentially, the students view every random variable as having a uniform probability distribution without consideration of the underlying random processes. The last way of thinking described is that students do not make a distinction between theoretical and empirical distributions. Students appeared to treat the empirical (frequency) distribution of the variable as being a copy of the theoretical distribution. One approach to addressing this way of thinking according to Magalhães was to have the students encounter a situation where values of the variable in the theoretical distribution were missing from the empirical data. Wild (2006) also mentions this distinction; he argues that the distinction lies in that we see the variation in the data (empirical) versus us imagining a potential model for the process that causes the variation (theoretical).

Prodromou (2012) tackles the distinction that Wild and Magalhães see between empirical and theoretical distributions in an alternative way. Prodromou described two epistemological perspectives for distribution; the first perspective is data-centric while the second centers on modeling and entails sample spaces and probability. Using a basketball simulation to serve as a bridge between the two perspectives, Prodromou (2012) described what connections two pairs of students made. The simulation consisted of a player who shoots a basketball at hoop. The students are able to set the values of various parameters (e.g., angle of the shot, and distance from the basket) and whether there is any random noise in the parameters. Once the students have set the parameter values, they can then have the player shoot the ball and the system will record whether the player makes the shot or not. For the following discussion, I focus on Prodromou's third task: two pairs of students were presented with a distribution of values for one of the variables (e.g., the distance between the player and the basket) that the computer would randomly select from. The students could manipulate the distribution by moving a slider (for the expected value) and arrows (for variance). The computer would then run the simulation and the students could observe the changes in the histogram on shots made/missed.

Neither pair of students suggested a random process underlying the third task. This is unsurprising as the students merely set the frequency for the 10 pre-set bins of the histogram representing the modeling perspective (i.e., the theoretical distribution). One pair (Anna and James) viewed the modeling distribution as the intended outcome and the empirical distribution as what actually happened. However, there are two interpretations for "intention" here: "a) the intention is simply an expression of the pre-programmed deterministic nature of the computer—at least in their experience, or b) intentions are reflected in the actions of a model builder" (Prodromou, 2012, p. 295). The other pair of students (Sarah and Nick) appeared to imbue the computer with human desire to shoot the ball from preferred locations. The level of preference stemmed from the frequencies the students programmed in the modeling distribution, but Sarah and Nick acknowledged that the computer could and would choose to shoot from places other than those most preferred. Prodromou (2012) called this view, when taken together with the view that the modeling distribution drives the data-centric distribution, "stochastic intentionality". When the intervention moved from the data-centric to the modeling perspective, Prodromou found that both pairs viewed the modeling distribution as the "target" of the data-centric distribution. In particular, the modeling distribution is "the 'future' outcome and the data-centric distribution displays the results of the present mechanism at play" (Prodromou, 2012, p. 296). Prodromou noted that as students continued to negotiate the two perspectives, that they encountered obstacles to their use of deterministic mechanisms. Prodromou argued that the students needed to develop quasi-causal explanations to account for the random mechanisms that underlie the player's shots. This was evidenced by Sarah and Nick's acknowledgement that the player (the computer) could and would choose to shoot from various distances from the hoop.

Bridging the arrangement and lens views of distribution. The Oxford English Dictionary (2015) entry for distribution includes notions such as spreading out, sharing, and dispersal. Watson (2009) highlighted that the meaning of "distribution" changes depending both on the inclusion of adjectives and the usage. Watson noted that

> Moore and McCabe (1993) progress to describe distribution in terms of variation, which they treat as an undefined term, and variable, which is 'any characteristic of a person or thing that can be expressed as a number' (p. 2): 'The pattern of variation of a variable is called its *distribution*. The distribution records the numerical values of the variable and how often each value occurs' (p. 6). (Watson, 2009, p. 32)

The Moore textbook has changed very little: "the distribution of a variable tells us what values it takes and how often it takes these values" (D. S. Moore, McCabe, & Craig, 2012, p. 2, 2015, p. 3). Neither Watson nor Moore state what they mean by the phrase "pattern of variation". Moore's meaning for distribution appears to blend both "distribution as arrangement" (values taken on and how often) with "distribution as a lens" and is heart of the largest category of how 45 statistics texts defined distribution. Watson (2009), studying students' (primary through grade nine) inclusion of variation [i.e., lack of sameness] and expectation in generating graphical representations, noted that many students think about distribution as referring to a collection of data values that can be displayed graphically; consistent with Bakker and Gravemeijer's (2004) first stage. Watson also found that students at all levels tended to give some indication of variation within their representations, with older students tending to give clearer indications of variation than younger students. For example, a Grade 3 student drew the picture on the left of (Figure 2) while a Grade 7 student drew the graph on the right when asked to draw a graph of the average daily maximum temperature for a year. Importantly, Watson did not use the term "distribution" with the students. Watson prompted the students during the interviews to either arrange physical cards, draw their own graphs, or comment on graphs made by other (research-created_ "students". Students' meanings distribution in Watson's study are therefore unclear. However, their choice of graphical representations and their accompanying remarks still provide insight into their meaning for variation in

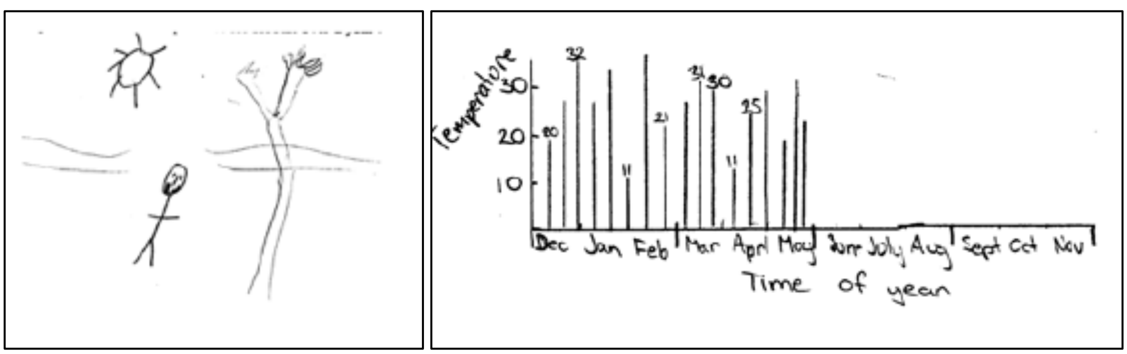


Figure 2. Two students graphs of the average daily maximum temperature for a year (Watson, 2009, pp. 42, 44).

the situations that Watson used. Watson (2009) hoped that encouraging students to a conception of variation, and to focus on it rather than focusing on measures of center, would lay a foundation upon which students could build an understanding of distribution.

Describing images of distributions. A second body of literature dealing with distribution focuses on students' thinking about describing visual aspects of dot plots, bar charts, or histograms. Authors often referred to these three types of visual display as "distributions". In many cases, the authors in this group referred to at least one of the authors in the prior section. In many introductory statistics textbooks there is an emphasis on the shape aspect of dot plots, bar charts, and histograms (Arnold & Pfannkuch, 2012). Arnold and Pfannkuck (2012) attempted to get 29 Year 10 students to describe dot plots and histograms (which Arnold and Pfannkuck continually refer to as "distributions") using the "language of shape". Students drew and organized the graphs under the headings "symmetrical", "sloped to the left", "sloped to the right", and "flat top". Eventually, the instructors introduced the conventional names of symmetric, right skewed, left skewed, and uniform. Graphs' shapes were taken as more than perceptual cues, meaning that students did not connect the shape of the graphs to any underlying process.

The language of shape that Arnold and Pfannkuck (2012) used encouraged students to engage in is what K. C. Moore and Thompson (2015) call static shape thinking. Drawing on perceptual cues, a student engages in static shape thinking when she views the graph as an object, much like a piece of wire bent into a particular shape and placed onto the plane. Figure 3 shows a set of what Arnold and Pfannkuck called "distributions", which they asked students to classify. Figure 3 comes from Lesson 3 in Arnold and Pfannkuck's study where the students "were given 15 contexts without graphs and asked to sketch the shape for these contexts with some possible values"

(Arnold & Pfannkuch, 2012, p. 5). They do not provide any examples nor say what these contexts consisted of. The students' sketches appear on the right side of each pair of images in Figure 3. After discussion on the students' sketches, the instructor provided the actual dot plots of the contexts (left side of each pair). None of the student-generated drawings have any axes or labels; their sketches literally look like wires in the plane. Arnold and Pfannkuck (2012) were not clear about what exactly the students thought their sketches represented; however Arnold and Prannkuck refer to the students' drawings as graphs as distributions.

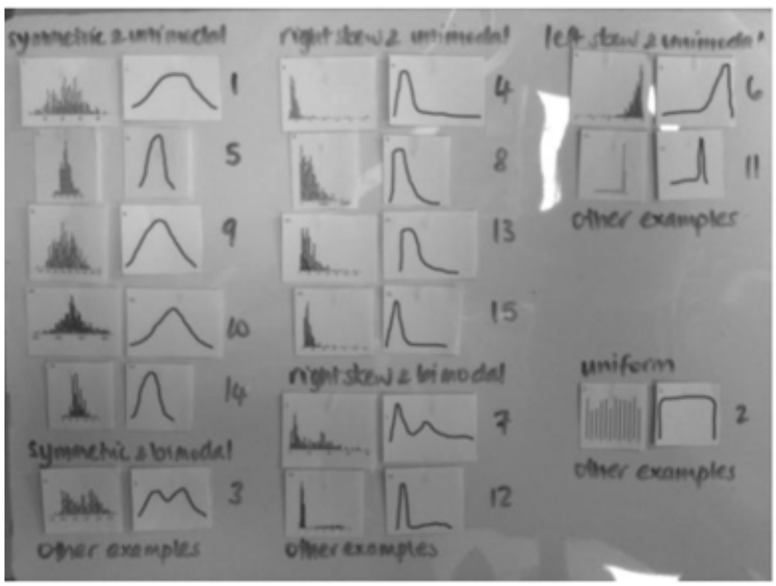


Figure 3. Collection of shapes which Arnold & Pfannkuck (2012) take as distributions (p. 5).

Arnold and Pfannkuck (2014) continued their prior work and wrote that "distribution is most often realized in a display such as a graph" and that students need to bring together ideas of center, spread, and shape (p. 1). They introduce their distribution framework, Figure 4, for analyzing students' descriptions of distributions. The italicized text indicates aspects of the distribution framework that are additions from the older "describing distributions framework" (Arnold & Pfannkuch, 2012, 2014). With the exception of specific features #1-3 and #6, the features are all perceptual ones. Using the distribution framework both to guide teachers in instruction and assessing students (Arnold & Pfannkuch, 2014), one possible messaged conveyed to students is that perceptual features are all that is important; that there is no need to attend to the underlying processes or quirks of the graphical system. Again, the idea of static shape thinking comes to bear as a possible result of Arnold and Pfannkuch's intervention.

Overarching statistical concepts	**Characteristics of distribution**	**Specific features measures/depictions/descriptors**
Contextual knowledge	Population	1. *Target population* (e.g. New Zealand year 5–10 students) 1. Other acceptable population (e.g. year 5–10 students)
	Variable	2. Variable 3. Units 4. *Values*
	Interpretation	5. *Statistical feature described in contextual setting* (e.g. interpreting right skew as very few high test scores, with most test scores between 20 and 50 points)
	Explanation	6. *Possible reason for a feature* (e.g. bimodal due to gender for kiwi data)
Distributional	Aggregate view	7. General shape sketched (correctly) 8. Hypothesis and prediction
	Symmetry	9. Overall shape
	Modality	10. *Modality*
	Skewness	11. Position of majority of the data (to the left or the right)
	Individual cases	12. *Highest and lowest values*
Graph Comprehension	Decoding visual shape	9. Overall shape 13. *Parts of the whole* (splitting the distribution into parts and describing the parts as well as the whole) 10. *Modality*
	Unusual features	14. Gaps 15. Outliers
Variability	Spread	16. Range, 17. interquartile range 18. *Range as an interval* (e.g. heights range from 132-197 cm) 19. *Interval for high and/or low values (*may be describing a tail) 20. *Interval for groups* (e.g. in the case of a bimodal distribution)
	Density	21. Clustering density 22. Majority (mostly, many) 23. Relative frequency
Signal and noise	Centre	24. Median, 25. Mean
	Modal clumps	26. *Peak(s)* 27. *(local mode)* 28. *Modal group(s)*

Figure 4. Distribution framework (Arnold & Pfannkuch, 2014, p. 2).

Distribution in service to another concept. Part of the research on individuals' understanding of distribution focuses not so much on distribution but rather another concept and how distribution plays a role. The most common examples include the concepts of variation and hypothesis testing. In the case of variation, there is a common mentality that "distributions are used to describe and model variability" (Peck, Gould, & Miller, 2013, p. 24). This approach is consistent with Wild's framing of distribution as a lens to explain variation. Lehrer and Schauble present a sequence of studies centered on students' reasoning about naturally occurring variation. In the first study, they framed distribution as an accounting of both the "true" measure and measurement error for some attribute (Lehrer & Schauble, 2002). Working with mostly the same students a year later, they aimed to have the students extend their understanding of distribution to include interpreting changes in the distributions and see distribution as structured variation (Lehrer & Schauble, 2004). Lehrer and Schauble (2002) got fourth grade students to begin thinking about variation in the context of measurement error by first having them each measure the height of a flagpole and then reason about the measures with displays of their own invention. Using the displays, the teacher guided the students into discussions about more and less trustworthy measurements and where the flagpole's true height was. Most students reasoned that the more measurements in the same bin there were indicated that those measures were trustworthy and that the true height would be more likely a member of that bin. Other students reasoned that measures in the middle region (including the most frequent bins) of their displays could be better trusted. These students put forth an argument that this region was a result of the measuring process. As class discussion progressed, students grappled with issues of precision and attributing error to various causes. In later follow up interviews, the researchers asked the student to construct the distributions for the measurements made by an imagined group of students of a statue's height using a more and a less precise tool. The interviewed students generated (roughly) symmetric displays in both cases, but tended to have the display for the less precise tool have a larger spread of values than the more precise tool.

The Lehrer and Schauble sequence of studies points to an aspect of variation that is difficult for students to distinguish: sources of variation. In the case of the rockets, the nose cone being rounded or pointed provided a natural source of differences in the launch heights of the rockets. The measurement tool used acted as a different source of variation in the values. Identifying different sources of variation can help students start to think about different types of variation: variation in the value of a single object's/living being's attribute over time, variation in the value of attribute as we shift our attention between different objects/living beings, and variation in the value of an attribute as we shift our attention from one collection to another (what Lehrer and Schauble (2004) refer to as sample-to-sample variability).

Conducting a hypothesis test requires using a distribution, particularly a sampling distribution. Saldanha and Thompson (2003, 2007) describe a multiplicative conception of sampling as entail imagery that includes viewing a sample as quasi-proportional to the parent population and the anticipation that when repeating the sampling process (i.e., the random process) there will be some variation in the results. Hatfield (2013) extended this multiplicative conception of sampling with his APOS-Sampling/Sampling Distribution framework. At the process level, Hatfield theorizes that as students repeat the sampling process, they will lump values of the statistic of interest together to generate what Saldanha and Thompson call a proto-distribution. The imagery involved here entails the student imagining the accumulation of outcomes from a random process. While in this case the distribution is a sampling distribution, (i.e., the distribution of a statistic for random samples of size *n*), the imagery is nearly the same as that of my target meaning for the distribution of a random variable.

Technical usages of distribution. Throughout much of the history of probability and statistics, the term "distribution" is synonymous with "probability distribution". While many mathematicians, statisticians, and logicians have tackled the issue of probability distribution, I will focus on the work of two men, R. von Mises and A. N. Kolmogorov. I choose to focus on these men for two reasons: 1) their work considerably impacted the development of statistics and probability both then and now, and 2) both describe what he meant by the phrase "distribution".

Von Mises, a champion of the frequentist perspective, viewed probability as the applied mathematics of mass phenomena (Weisberg, 2014). For von Mises, the distribution of a random variable refers to relative frequency (probability) for each value of the attribute within the collective (von Mises, 1981). He makes a distinction that for discrete random variables, there are a finite number of fractions representing probability values but in the continuous case, the fractions represent the probability density per unit length. He makes use of the following metaphor to introduce his notion of distribution: Imagine that you are tasked with doling out one kilogram of mass in a straight line, one meter in length. If you put the same amount of mass at every place along the line, you essentially have a rod of uniform thickness; if you make amounts uneven along the line, making thinner and thicker spots, you have a rod of variable thickness. In any case, you made tangible the distribution of mass along the line. That is to say that you can give a measure of the mass or the mass density (per unit length) at each and every point along the line (von Mises, 1981). The metaphors of mass and density carry though today when we talk about distributions; probability mass functions are for discrete random variables and probability density functions are for continuous random variables. Describing a distribution requires that a person coordinate his/her location along the line (value of the random variable) along with the measure of the mass/mass density at that location (probability/probability density). This last statement is similar to how Leavy defined distribution; "the arrangement of values of a variable along a scale of measurement resulting in a representation of the observed or theoretical frequency of an event" (2006, p. 90). However, there are a couple of important distinctions. Leavy's statement defines the concept of distribution while von Mises's metaphor centers on describing a distribution, not defining one. A second distinction is von Mises's mantra "first the collective, then the probability" that highlights the centrality of two central concepts for distribution: collective and mass phenomena (von Mises, 1981, p. 24; Weisberg, 2014, p. 233). For von Mises, mass phenomena are any events/processes that we can imagine being repeated an unlimited number of times while anticipating that each repetition will not be necessarily identical. If you were to carry out a mass phenomenon (in practice or imagination), and record the results, you construct what von Mises calls a collective (von

Mises, 1981). Thus, when defining distribution, we cannot leave out the role of the random process as the collective's generator. While Leavy's definition echoes von Mises's metaphor, the role of a random process is absent.

Kolmogorov (2013) laid out his own axiomatic foundation for probability based on Lebesque's theories of measure. For Kolmogorov, distribution related most strongly with the notion of a cumulative density function. In his treatise, he defined the distribution function of a random variable, *x*, as $F^{(x)}(a) = \mathrm{P}^{(x)}\left(-\infty, a\right) = \mathrm{P}\{x < a\}$. This is the say that the output of the distribution function for the random variable *x* at the value *a* is the output of the probability function, $\mathrm{P}^{(x)}$, for the set $(-\infty, a)$. Kolmogorov defines P(*A*) as the summation of all p_i, the probability value given to element ξ_i, where ξ_i belongs to the set *A* (Kolmogorov, 2013). More clearly put, the distribution function of a random variable gives the cumulative probability of observing values of the random variable less than the input value *a*. This differs from modern definitions of the cumulative density function in that the later provides the cumulative probability of observing values less than or equal to *a*. Much like von Mises, Kolmogorov's approach to distribution contains the joining of two features: the value of the random variable and the cumulative probability of the random variable up to the given value. While Reading and Canada's (2011) refinement of Leavy's definition includes probability, their definition of distribution stops short of what Kolmogorov meant. Their notion of distribution pairs the variable's value with the probability of being at that value. Kolmogorov explicitly linked the value of the variable with the accumulation of probability, making distribution a more dynamic idea.

Comparing different meanings for distribution. Reading and Canada (2011) point out that many statistics education researchers believe that understanding distribution depends upon understanding features of distributions. However, this way of understanding distributions is rife with incoherence when we step back and reason quantitatively. Reading and Canada's list of what distribution depends upon serves as a list of attributes. Thus, students must first understand measures of center, spread, and shape in order to understand distribution. This is logically equivalent to saying that students need to first understand a person's height before they can understand/imagine a person. If random variables' distributions do depend on the a priori existence of the measure of these attributes, then what understanding can students develop for distributions that do not have the "standard" attributes students believe are necessary for distributions to exist? For example, how should students come to develop the idea of distribution for categorical stochastic variables that do not possess numerical values with which the students can calculate the value of the sample arithmetic mean? Further, how should students define the Cauchy distribution when this distribution does not possess the attributes that they have come to believe are the foundation of the distribution concept?

As I've already written, there are distinctions between how von Mises and Kolmogorov defined distribution and how statistics education researchers explain the concept. The "distribution as arrangement" view leaves the random process out of the formation of a distribution and makes the concept static. This view of distribution supports two images; distributions are pictures (i.e., histograms, probability density curves) and distributions are data collections. These two images are rather complimentary especially when the picture is a histogram or dot plot. In most cases, the treatment of distribution as data collection is implicit, but some authors make this explicit. Peck, Gould, and Miller (2013) write "research suggests that making the transition from talking about individual values to considering properties of the entire collection—the distribution—is a key conceptual leap" (p. 25). In many works (e.g., see Arnold & Pfannkuch, 2012, 2014; Ben-Zvi, 2004; Leavy, 2006; Lehrer & Schauble, 2002) replacing "distribution" with "collection" preserves the author's message and potentially makes the message clearer. I do not know what Peck, Gould, and Miller mean by "collection"; however, their usage appears consistent with considering a collection of data values. When I use the word "collection", I do not mean a set of numbers that I look at. Rather I mean that I have a set of observed values (numerical or not) for one or more attributes for each object/living being who is a member of some population and the set has 1) structure, 2) properties based upon the members of the set, and 3) has context. Students thinking about and reasoning with collections is important for data analysis as such thinking undergirds exploratory data analysis and lays the foundation for confirmatory data analysis. Treating distribution as synonymous with (data) collection underpins Bakker and Gravemeijer's (2004) three-stage framework. Thus, if all that statistics education researchers mean by "distribution" is "collection", then why introduce a new term? The concept of distribution did more work for both von Mises and Kolmogorov than what thinking of distribution as an arrangement/distribution as a collection allows. For von Mises, the notion of the collective is better parallel to what today's statistics education researchers refer to as distribution.

Thinking about distribution as telling us not only how often a random variable takes on a particular value but also what values the variable represents highlights a confounding of the notions of distribution and domain. The domain of any variable tells us what values are admissible for that variable; no need to invoke the distribution concept. I suspect that the confounding has roots in modern statistical language, particularly phrases such as "discrete distribution", "continuous distribution", "joint distribution", and "empirical distribution". The adjectives "discrete" and "continuous" do not operate in the same manner as the adjectives "joint", "marginal", and "conditional", nor like the adjective "empirical" or noun "sampling". "Continuous" and "discrete" (which includes the adjective "finite") do not modify the distribution concept, rather they are modifying the *domain of the random variable*. This stands in contrast to the terms "joint", "marginal", and "conditional" which indicate which specific type of behavior is being described. The adjective "empirical" appears to stand-alone and is most often a reference to examining the empirical distribution function for kernel estimation and testing the goodness-of-fit of proposed distributions. The term "sampling distribution" is a shorthand way of invoking an extension of the distribution concept: the distribution of [some statistic] for samples of size *n*". The subtly of the differences between domain modifiers (i.e., "discrete", "finite", and "continuous") and behavior modifiers (i.e., "joint", "marginal", and "conditional") as well as "empirical" is not well covered in any statistics text. Any investigation into how people think about the distribution concept must take these subtle distinctions into account.

Distribution as an arrangement of values also appears in the literature through examination of probability density curves. However, as mentioned previously, students focus on perceptual features of these graphs and could be conflating the graph (a representation) with the distribution (that which the graph represents). In their development of distribution, neither von Mises nor Kolmogorov used a graph of a distribution function; the imagery they used centered more on the building up of the collective (von Mises) and the accumulation of probability (Kolmogorov). Overemphasis and reliance on graphs of probability density functions poses some hazards to students understanding of distribution. For example, consider the two probability density functions whose graphs appear in Figure 5.

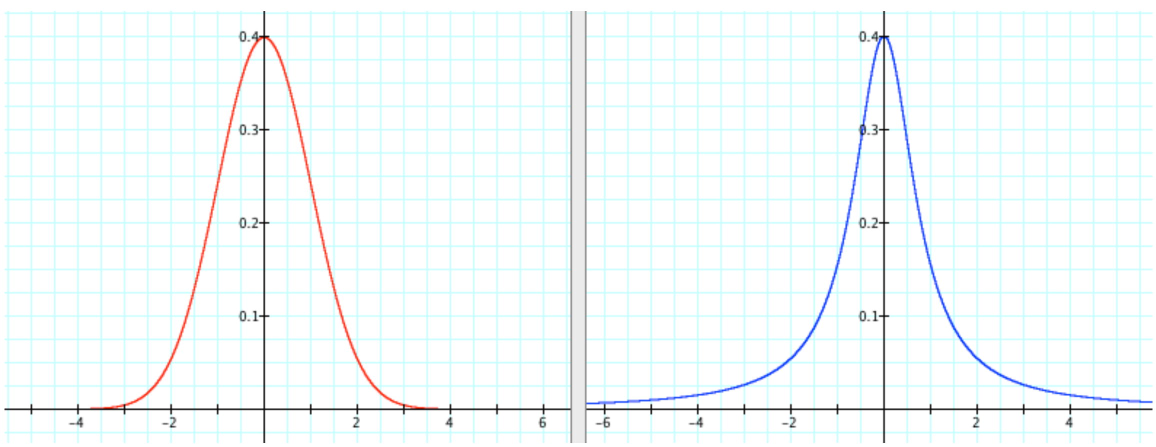


Figure 5. The graphs of two PDFs.

Using Arnold and Pfannkuch's (2014) framework, we could propose what a high level descriptions of the two graph might consist of: For some population of people (e.g., New Zealand year 5-10 students) both random variables seem to have all real numbers as values, both are symmetric (no skew), with a majority of the values in one central modal clump. There are no gaps or outliers. Both graphs appear to be centered at zero thus the values of the median and the sample arithmetic mean are zero. The graph on the right has less variation around zero (smaller interquartile range) than the left graph, but the left graph has less variation over all compared to the right graph (the left graph "hits" zero at ± 4 as opposed the right graph which does not). We could offer up many refinements to these descriptions, but I want to highlight a couple of features. First, only when I started writing about variation did the descriptions become specific to one graph verses the other. Second, in using the aforementioned framework, I was able to hit every category without ever talking about probability or a random process. This approach brings a question to my mind: if a student taught to think of distributions using this framework and who has learned some named distributions (e.g., Normal, Binomial, etc.) examines these two graphs, would he/she classify the distributions as being the same up to value of parameters or completely different?

However, these two graphs are not of the same distribution; the graph on the left is that of a standard normal distribution while the graph on the right is of a Cauchy distribution (which has no moments higher than the zeroth moment). Distribution as arrangement appears to hide essential aspects of the distribution concept from students; specifically, the role that a random process plays in the distribution.

Treating distribution as a lens by which we look at variation is also at odds with the notion as conceived by von Mises and Kolmogorov. Bakker and Gravemeijer (2004) wrote "without variation, there is no distribution" (p. 149) while Watson (2009) claimed "variation in data create distributions" (p. 34). Both of these statements echo Wild's (2006) definition that a variable's distribution is "the pattern of variation in a variable" (p. 11) but create a stronger relationship between variation and distribution. Consider an attribute for which every member of the population has the same value of, for example, the "humanness" of humans. No matter which member of population you examine, every member has the value "human" for this attribute. There is no variation in this attribute, thus, according to Bakker and Gravemeijer as well as Watson, there can be no distribution for this variable. At least Wild's statement allows for a constant pattern of the sameness. Kolmogorov's and von Mises's concepts of distribution do not depend on there being any variation in the value of the random variable.

None of the statistics education articles listed in the prior sections referenced either von Mises or Kolmogorov, nor any other treatise on the development of statistics; the closest that any author gets is a reference to some edition of the D. S. Moore *Introduction to the Practice of Statistics* textbook. While I do not believe that citing original statistical treatises should be a requirement of statistics education articles, the lack of such citing might explain the apparent schism between the meaning of distribution laid out by statistics education researchers and the meanings developed by two forebears of Statistics. Viewing distribution as a static arrangement of values or "the pattern of variation in a variable" (Wild, 2006, p. 11) stand in contrast to both von Mises's and Kolmogorov's notions of distribution. While von Mises's usage of distribution appears similar to distribution as arrangement, there are critical distinctions. Von Mises's usage depends upon thinking about a random process and the construction of a collective from that process before the naming of probability values. Both distribution as arrangement and distribution as lens appear to leave an underlying random process out of the conception of distribution. While Kolmogorov did not stress the random process as much as von Mises, he did make central the notion of accumulation that statistics education researchers leave out altogether. The role of imagery is also different. While von Mises uses a visual metaphor, the notion of arranging values is more in line with coordinating two aspects than placement in a pictorial way, typical of the statistics education research cited. Statistics education researchers jump straight to graphical representations (e.g., histograms, graphs of probability density functions) and appear to emphasize that these representations are distributions. Von Mises's usage of imagery focused on metaphors to convey connections between key ideas for distribution. Kolmogorov's meaning for distribution does not rely on imagery; he presents his development of distribution in the language of algebra, metric spaces, and calculus. Thus, I propose with this dissertation study to investigate the meanings for distribution that students develop when instruction seeks to support a new meaning for distribution. I seek to promote thinking about a random variable's distribution as *the accumulation of a random process' outcomes with respect to the random variable's value*. The mental imagery underlying this way of thinking entails imagining that as we move through sequentially through the values of the random value, we continually add on more outcomes until we reach the point where we have obtained all possible outcomes. This meaning for distribution keeps Kolmogorov's notion of accumulation and von Mises' emphasis on random processes central while dispensing with focusing exclusively on graphical arrangements. I provide a conceptual analysis of this way of thinking in Chapter 4.

Randomness and Random Process

One of the foundational ideas of distribution is the notion of randomness. However, "randomness" resists definition and most authors tend to use the term as an adjective, focusing instead on what "random" modifies (Batanero, Green, & Serrano, 1998). This being said, researchers in psychology, decision-making, as well as mathematics and statistics education have investigated individuals' conceptions of randomness since the 1950s. Given that the focus of this dissertation is the concept of distribution, I will not attempt to convey the entirety of this body of research. One feature that makes randomness difficult is that "random" lives in two worlds: the technical and the everyday. This dual life gives "random" lexical ambiguity (Kaplan, Fisher, & Rogness, 2009; Kaplan, Rogness, & Fisher, 2014). Investigating "randomness" lexical ambiguity Kaplan et al. (2009) found that twenty-nine of sixty-one students defined "randomness" along the lines of "haphazard" and "unplanned". Another twenty-seven students treated "random" as being about selecting without criteria, (a known) order/pattern, or bias. The authors juxtapose these students' definitions with that of the normative definition of random event; an event "for which no one outcome can be predicted, but there is knowledge of the long-term distribution of the outcomes" (Kaplan et al., 2009, p. 13). In a subsequent study, Kaplan et al. (2014) replicated their activity with another set of students. However, in this second study, they collected students' definitions after intervention as part of the course final as opposed to before the students had any instruction. Kaplan et al. (2014) used six coding categories for students responses; incorrect (e.g., "unknown"), by chance, without order/reason, unexpected/unpredictable, without bias, and equally likely. The ordering of the categories reflects the authors' view of category being closer to the statistically sound meaning for randomness. While in the first study, 47.5% of the students defined random as haphazard, unsurprisingly only 6% of the students in the second study did so. Most of the students in the second study, 40%, gave a definition consistent with viewing randomness as equally likely (the authors' highest category). This category equates "randomness" with equi-probability (chance), which is a common historical perspective (Bennett, 1993). However, this meaning muddles the distinctions between randomness and probability, creating unnecessary lexical ambiguity.

Batanero and Serrano (1999) found that secondary students use a mixture of arguments to justify whether or not something is random. Their categories include the students noticing 1) a regular pattern, 2) the absence of a regular pattern, 3) result frequencies are similar, 4) result frequencies are dissimilar, 5) there are runs, 6) there no runs, and 7) un-predictability. With the exception of the last category, the categories come in polar opposites. These categories are similar to those of Kaplan et al. (2014) with (1) and (2) linking with "without order/reason", (3) with "equally likely", and (7) with "unexpected/unpredictable". The lack of a pattern links to the notion of complexity described by Falk and Konold (1994) and championed by Kolmogorov. Here, an individual's difficulty in condensing a sequence is a measure of how random the individual believes the sequence to be; the easier time she has compressing the sequence, the less random she believes the sequence to be. This notion of perceived randomness has roots in the heuristics that individuals use in making judgments (Kahneman & Tversky, 1974). Examining result frequencies harkens back to von Mises's adherence of the impossibility of a gambling system (von Mises, 1981). Here, the individual would attempt to find a system by which the relative frequencies of each result can be altered (e.g., picking every third result). The presence/absence of runs highlights students not viewing individual results as independent of other results.

Defining "random" is a difficult and complex task (Falk & Konold, 1994, 1997). As such, some authors argue that instead of worrying about how to define randomness, we instead focus on random processes. Wagenaar (1991) argues that viewing randomness as a property of a sequence instead of the underlying generative process is problematic given people's reliance on heuristics. Falk (1991) argues that this approach is preferable as "random process" has a more agreed upon and stable definition than randomness. Both Wagenaar and Falk point out three characteristics of random processes: 1) there is a fixed outcome space, 2) there is a selection process that is unbiased, and 3) all iterations of the process are pairwise independent (i.e., the outcome of any trial bears no impact on the outcome of any other trial—except when sampling without replacement).

Konold, Harradine, and Kazak (2007) report an investigation to help students understand that the most frequently occurring values of a random variable are the result of there being more ways to get those values. As part of their study, the authors asked students to create a "data factory" in TinkerPlots. In doing this, students needed to conceive of objects and living beings as having attributes that we can classify and measure in order to create data. The students needed to pick a type of object/being and then select some attributes to use in their data factories. After this point, students need to come up with possible values; in essence, they needed to fix the outcome space. Students used these values for each attribute and used a built-in tool such as a spinner to create a selection method. In constructing their tools, students had the option of adjusting the underlying distribution. For example, they could make a spinner have equal sized wedges for each value (uniform) or they could make different values have different sized wedges. Students could run their factories, examine the data generated and then revise their factory's elements. This data factory approach to random processes seems in line with Liu and Thompson's (2002) push to focus on the mental imagery and operations that would enable a student to construct a coherent understanding of stochastic situations. They argue against focusing on the ontology of randomness, random sequence, and random process and instead focus on what is meant. One set of imagery is that of a process that does not have well-defined inputs but still generates data while being simultaneously predictable in the long run and unpredictable in the short run.

I must address a discrepancy between the description of random processes described thus far and how a sizeable number of statisticians think about the phrase. The random processes described thus far are also called "stochastic processes" and are most often juxtaposed with deterministic processes. When I use the phrases "random process" or "stochastic process" I mean a process an individual envisions as taking loosely defined inputs, and through a fuzzy rule (ill-defined) returns a datum; the individual sees the process as infinitely repeatable but the results of individual trials are not the always the same (nor always different) and cannot be correctly anticipated with any regularity (i.e., the processes is replicable but not reproducible). My meaning is in line with what von

Mises described as constituting "mass phenomena" (von Mises, 1981). Kolmogorov (2013) drew on von Mises's work when he described the generation of data to provide a pragmatic grounding for his theory of probability:

> 1) There is assumed a complex of conditions, $\mathfrak{S}$, which allows any number of repetitions. 2) We study a definite set of events which could take place as a result of the establishment of the conditions $\mathfrak{S}$. In individual cases where the conditions are realized, the events occur, generally, in different ways. (Kolmogorov, 2013, p. 3)

Kolmogorov's $\mathfrak{S}$ refers to the loosely defined inputs that Liu and Thompson (2002) and I refer to in our meanings for stochastic (random) process. The realization of those conditions is the fuzzy rule of the process.

This usage of stochastic process stands in contrast with the more common definition: a stochastic process $\{X(t), t \in T\}$ is a collection of random variables with parameter space T (see Beichelt & Fatti, 2002; Ross, 2010). Where many statistics texts go from this definition is to treat "stochastic process" as a synonym with "time series". This usage, while useful in certain contexts, overly narrows the notion of a stochastic process to only admit processes that look at the value of a single attribute for a single object; for example, the number of pizzas sold by a particular store each Friday or the diameter along a specific length of wire. My meaning for stochastic process certainly admits these examples; however, consider the case of the height of individuals 20 years or older, living in the United States. In this case, we have one attribute (height) but multiple objects; a time series is built to embrace and use the dependency between consecutive states (i.e., $X(t_1)$, $X(t_2)$, $X(t_3)$, etc.) which does not necessarily exist in the heights example. Given that students' exposure to Statistics does not begin with time series analysis, using a meaning for stochastic process that encompasses more phenomenon that statisticians deal with will help these students generate better models for where data come from.

Related to the present discussion of stochastic process is the concept of a "second-order stochastic process". Typically, this phrase refers to a time series where we assume the existence of second moments. When I use the phrase "second-order stochastic process" I will mean a stochastic process (my meaning, not time series) that involves one or more stochastic processes as sub-components. For example, we may think about getting the height of a woman who is 20 years or older, living in the United States, as the realization of some [first-order] stochastic process. If we repeat this process for a total of 30 measures and then use a second first-order stochastic process to get 30 heights of men, we can then find the value of a difference statistic; say the difference in values of the sample arithmetic mean for the two groups. The value of the difference statistic is one realization of the second-order stochastic process. To get a second realization, we must carry out each of the first-order stochastic processes again (30 times each) and then calculate the value of the difference statistic. In my framing, a first-order stochastic process leads to describing the distribution of an attribute shared by many objects/living beings while a second-order stochastic process leads to describing the sampling distribution of some statistic/estimator.

Saldanha (2016) discusses students' difficulties in conceptualizing "stochastic experiments" (i.e., what I would call stochastic processes). Students were asked to decide whether or not a movie ticket-taker seeing at least two people he knows is an unusual event given the ticket-taker knows 300 of 30,000 people and the theatre holds 250 people. Saldanha found that students stumbled in their conceptualization of stochastic processes as they wrestled with choosing which assumptions they needed to make, envisioning the population and the sample, and what constitutes a trial (a single run of the stochastic process). Introducing yet another name from stochastic process, Kuzmak (2016) proposes what a mature schema for understanding common random phenomena consists of: a mechanism and a way to repeat activating that mechanism, outcome sequences, and predictability (Figure 6). The schema has three main categories (numbered) with several characteristics in each category (lettered). Kuzmak reports on 24 college students' interactions with a machine that shakes a tray holding equal numbers of red, yellow, and blue marbles. The students predict which marble falls out of the tray, keeping track of the color and whether or not they correctly guessed for 12 trials. Of the 24 undergraduates, only four gave responses that Kuzmak classified as fitting the mature schema for random phenomenon. Specifically, Kuzmak argues that these four students discussed that the expected number of successes out of 12 trials was 4, this number was what others should expect even though the subjects had higher success, that strategies for prediction don't matter, and that marble shaking machine phenomena was random.

A common random phenomenon has:	
1. A physical mechanism, with a method to run repeated trials that each produce an outcome:	a. The *mechanism* has *features* that ensure no bias in favor of any particular outcome b. There is *a set of possible outcomes* for each trial, that set numbering more than one (=*n*); and each possible outcome has equivalent possibility, equal potential, equal *probability* (=*1/n*) to occur on each trial c. Outcomes on successive trials are independent, generated by the same mechanism, which is stable over time
2. Outcome sequences:	a. Over the long run (*m* trials), each of the possible outcomes has equal *expected frequency* in the outcome sequence (=*m/n*) b. There is *variation* in the frequency and pattern of occurrence of the possible outcomes among outcome sequences c. Over the long run, outcome sequences show no systematic order or pattern, and are usually mixed-up looking d. The probability of the next outcome in a sequence is independent of past outcomes, even when there has been an unusual sequence of outcomes such as a long streak of a single outcome category e. Orderly/patterned sequences are possible to occur by chance as *rare events*, as they are in the *set of all possible outcome sequences*
3. Predictability (by self/ others):	a. Don't know which outcome will occur, it could be any of the possible outcomes; so difficult to predict b. By chance, no matter which outcome one predicts, one has *probability* of prediction success *=1/n* c. Over the long run (*m* trials), *expected prediction success* is *m/n* times, or *1/n* of the time; and expect *variation* in prediction success across trials d. Long streaks of prediction success or failure are possible by chance as *rare events*, because such events are in the *set of all possible prediction results* e. Particular prediction strategies are irrelevant to prediction success f. "Being lucky" is not a causal influence on prediction success

Figure 6. A mature schema for random phenomenon (Kuzmak, 2016, p. 182).

While Kuzmak's work is a step in the right direction, I must point out the shortcomings within this schema. First, the given schema entangles not only the ideas of a first-order stochastic process, but also second-order stochastic process, the prediction concept, as well as the idea of a sampling distribution for the number of correct guesses out of 12 (i.e., a binomial distribution of prediction success).

Second, the first category (physical mechanism) is the first-order stochastic process but makes the mistake that equiprobability of outcomes is a necessary condition for mature understandings. As von Mises and even Laplace noted, the notion that all outcomes have the same probability of occurring should be the first thing abandoned when an individual works with probabilistic notions in the real world. Kuzmak grounds the process in some physical mechanism that produces independent trials. While there are stochastic processes that produce independent trials, there are also those that do not; time series being an important example. Kuzmak's mature schema disavows for time series to be consider random phenomena. The outcome sequences that Kuzmak proposes as the second category of her mature schema of random phenomena is better described as a schema for students' understanding of a binomial situation; that is, a second-order stochastic process where a first-order stochastic process is repeated a certain number of times and the number of "successes" in that number of trials is the outcome of interest. Students must be able to coordinate the outcome of the first-order process (the falling of a marble and observing the color), predicting the color of the falling marble, and the outcome of the second-order process (the number of successful predictions out of 12 attempts). The mature schema as described only works in specific binomial contexts. I suspect that if we were to take these same students and present them with different trays that have different numbers of each color of marble (l red, m blue, and n yellow where $l \neq m \neq n$), only a few of their answers would change. If we were to change the context completely, say to predict the heights of people, their answers could change more dramatically. In either case, this mature schema would be of little use to the students. Kuzmak's third category focuses on the role that prediction plays in stochastic processes; primarily that prediction success is not guaranteed. Again, the equal probability bias is a central component of this category appearing in characteristics 3b and 3c. Sadly, while the experimenter asked three of the four the students showing the mature schema what "random" meant to them (as well as the other 20 students), Kuzmak (2016) does not share any student's response to this question. Instead, Kuzmak shares students' answers to the question of whether or not they believe that the phenomenon is random. Table 1 shows the responses of eight students to this question as reported by Kuzmak. Students A, B, and D are purported to have a mature schema for random phenomenon while the other students have various immature understandings. In terms of a stochastic process, Kuzmak coded Students A, B, D, and G as seeing the marble shaking trays as being random while Students N, R, and V do not. For these three students Kuzmak contends that they "fail to show knowledge that prediction strategy has no influence on prediction success for random phenomena because each outcomes is equally likely" (Kuzmak, 2016, p. 192). Student W is coded as having abandoned or expressing doubt about whether the marble shaking experiment is random.

The responses of Students A, B, D, and G focus almost exclusively on the uniformity of the marbles; only Student D mentions the tray. The other students all mentioned aspects beyond just the marbles. Of particular note is Student W's response. I would characterize W's response as expressing the fuzzy rule nature of the phenomenon. Student W does not make a claim about the randomness of the phenomenon but rather notes that he/she is missing information and wants to do more experimentation.

Table 1. Students' Responses to Whether the Marble Trays Are Random

A: They [the trays/phenomena] all of have the same amount of marbles, and all the marbles are the same. Then they're random. (p. 187)
B: On the basis of the fact that the trays seemed to be shaped identically, but more important, that they seem to contain the same number of evenly shaped, evenly weighted marbles, yes, I'd say they're random. (p. 187)
D: Not totally…pretty random. If they're the same number of marbles, and if they're the same size and weight, if the trays aren't tilted…I don't know if the numbers are large enough, but they appear to be pretty random. (p. 187)
G: I would say so. Assuming all the balls are the same amount. They're all weighted the same also. (p. 190)
N: Yes, I would say so. Exptr: In other words, so in what sense are they [trays] random then? N: That the balls could come down any way given an infinite number of tests you did…" (p. 191)
R: Well, in the sense that it's too difficult to figure out which marble's gonna drop through, then it's random. But in the sense that, if you repeated the experiment the exact same way with the marbles sort of like distributed before you started, and the machine happened to work in the same way, again, then you'd get the same marble, so that wouldn't be random. But since it's beyond, you know, it is beyond possibility for repeating exactly, then you could call it random. (pp. 191-192)
Exptr: Could you describe these trays here as random? V: The trays with marbles in them? Exptr: Right. V: Yeah, I think so. Although, actually all the colors are—tend to bunch together a little bit. See? But, I think, yeah, they're probably random. The reds and the blues are much more than the yellows, bunching together… Exptr: OK. So, would you say that it was random the way the marbles came out of the hole? V: Not really sure. (laugh) I'm really not sure. They do tend to come out in pairs. Exptr: OK. Uh, and why would they come out in pairs? V: Because the colors tend to be bunched together. (p. 192)
Exptr: What do you mean by the distribution within the tray? W: Well, I mean, I would, I mean I don't know too much about physics (laugh). But I would think that if, if we had a lot of yellow marbles maybe bunched up together or something, closer to the center, that might influence the results… Exptr: [Later] Could you say that these trays here were random? W: Not conclusively. Exptr: And what do you mean by that? W: (pause) I haven't been allowed to have; I don't know all the properties of at work. I mean, I haven't had, I don't have enough trials with any of the trays to say anything. Nor do I know if the marbles are of equal weight or equal anything (laugh). Or if the trays are the same. (pp. 192-193)

All quotations come from Kuzmak (2016) with page numbers given.

There are two issues at hand with these responses. First, the meanings these students have for “random” within the marble shaking experiment context are highly problematic. Even the students identified as having mature schemas based their judgment of randomness on features of the marbles rather than the stochastic process. Student R’s response is interesting in that he/she appears to invoke that there are two stochastic processes at hand: the dropping of a single marble and a full set of 12 Bernoulli Trials. In both cases, R appears to equate “random” with “unpredictability”. The second issue at hand is much more problematic: the students were not actually asked whether the phenomenon was random, rather the experimenter asked whether the *trays* were random. Student F (not shown) directly asked the experimenter if the question was about the trays or the methods, while Student V asked for confirmation. The experimenter’s question could be why students focused so much on the marbles rather than on the stochastic process.

Probability

Probability is the engine that makes inferential statistics run. In particular, probability is the result of centuries of work towards one goal: the quantification of uncertainty. Since before the 1600s mathematicians, philosophers, logicians, and statisticians have attempted to resolve questions where uncertain outcomes dominate (Weisberg, 2014). Over the course of history, many scholars have engaged in what Thompson (2011) calls quantitative reasoning and quantification. In settling what a measure of uncertainty means (along with how to get a measure and what is meant by measuring uncertainty), scholars have taken different paths and arrived at their own meanings for the same notion, probability. Laplace considered the ratio of the number of desired outcomes to the number of all possible outcomes under the assumption of “equally likely” outcomes (Weisberg, 2014). Von Mises (1981) considered repeating some process indefinitely to build a collective and that the limit of the relative frequency of an event of interest was the probability of that event. Kolmogorov’s (2013) axiomatic, measure-theoretic approach has become the gold standard for probability theory. De Finetti (1974) and Savage (1972) regarded probability as dealing with measuring the amount of belief that an individual had for a particular outcome’s occurrence that they called “subjective” or “personalistic” probability.

Few practitioners of statistics will disagree with the analogy that probability is an engine. Regardless of which school of probability you ask (i.e., Frequentist, Bayesian, Conditional Frequentist, etc.) each acknowledges that the central ideas of probability allow us to move beyond merely describing a data set to using the data set as evidence for supporting or refuting claims. The members of these schools of thought have already carried out the quantification of uncertainty, something that students have yet to undertake. How practitioners think is often vastly different from how students think before, during, and after instruction.

Kahneman and Tversky (1974, 1982) described how individuals will use different heuristics when making judgments under uncertainty. For example, how representative an event (sample) is to the parent process (population) can influence a person’s estimate of the probability of the event. Another heuristic that they found that people use to measure uncertainty centers on the ease (or lack of) with which a person can imagine the event occurring; the more “available” an event is for the person to imagine, the larger the probability (the less uncertainty) there is for that event. Konold (1989) found that for some individuals, their way of thinking about probability did not match the use of heuristics nor was their thinking consistent with the schools of probability. Rather, these individuals appeared to view the goal of uncertainty to be the prediction of the next result; Konold referred to this way of thinking as the outcome approach to probability. Students also have a tendency to view events as equally probable when they do not perceive the many ways a compound event might occur (Lecoutre, Durand, & Cordier, 1990). Lecoutre et al. found that students and adults view the event of getting a five and a six as having the same probability as getting two sixes when rolling two dice. They hypothesized that not recognizing that event of (5, 6) is comprised of two smaller events. This way of thinking across multiple events is what they referred to as the equiprobability bias. Fielding-Wells (2014) found that when trying to pick the best card for playing addition-bingo, Year 3 students (7-8 years old) operated as though all of the sums of the numbers 1 to 10 were equally probable. Hatfield (2016a) found that 89 out of 114 undergraduate students (78%) conveyed a circular meaning for probability when asked to explain probability. Students would say that probability was the "chance" or "likelihood" of some event occurring. This finding is in line with Kaplan et al.'s (2009) findings regarding lexical ambiguity; students appear to use colloquial meanings as their dominant meaning for the technical term probability. What is perhaps most disturbing is that Hatfield collected the data *after* the students had received instruction and been tested on probability in their introductory statistics course. This suggests that 1) students' colloquial meanings are resistant to change, or 2) instruction did not challenge the students' colloquial meanings or make those meanings problematic.

Saldanha and Liu (2014) reviewed much of the literature on students' understandings of probability and proposed that a key conceptual scheme for understanding the measurement of uncertainty is a stochastic conception. They define a stochastic conception as "a conception of probability that is built on the concepts of random process and distribution" (p. 393). They argue that in the quest to support students developing coherent probabilistic reasoning, instructors need to conceive of probability as ways of thinking rather than skills and design curriculum that supports this. Fielding-Wells (2014) had students play addition-bingo and track results, thereby allowing students the opportunity confront their equiprobability bias. Getting students to think about a random process can emerge from having students construct data factories (Konold et al., 2007) as well as using a simulation approach. Students need to be able to view a trial as an iteration of the random process, that the observed value(s) of this trial make up the outcome (Horvath & Lehrer, 1998). Individuals of all ages and backgrounds struggle just as mathematicians, statisticians have to construct a meaning for probability. Liu and Thompson (2007) investigated eight in-service teachers' understanding of probability. They generated a set of models for the teachers' meanings for probability shown in Figure 7.

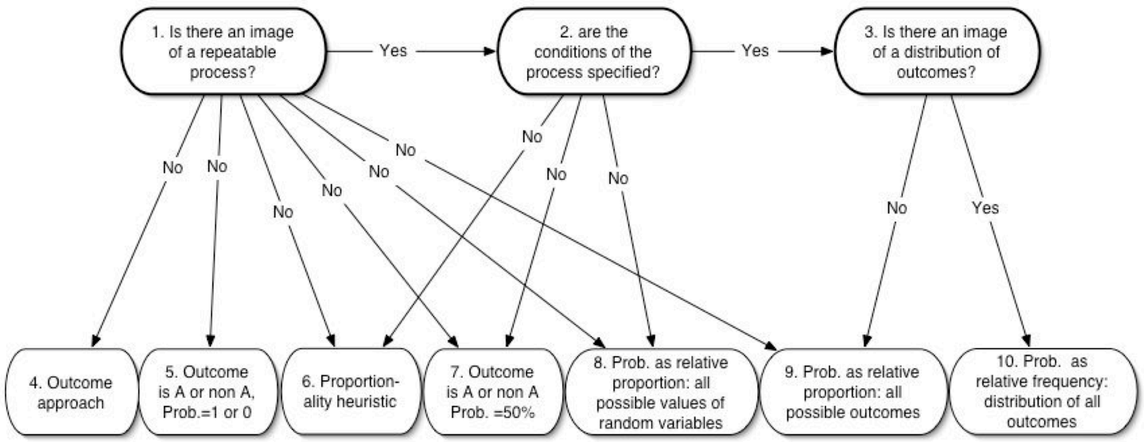


Figure 7. Meanings for Probability (Liu & Thompson, 2007, p. 124)

The three questions at the top of Figure 7 reflect a way of thinking that underpins conceiving a situation stochastically. The lower seven bubbles each reflect seven different meanings for probability based upon answering the three questions. Liu and Thompson only classify viewing probability as relative frequency (from a distribution; #10) as being a stochastic conception of probability. While teachers mostly started out with a non-stochastic conception of probability, during the course of the seminar, the teachers began to conceive of more situations stochastically and in several instances, recognize that the same situation could be conceived both stochastically and non-stochastically. For probability to do the work necessary for distribution (and statistical inference), individuals need to be able to conceive of situations stochastically.

Cobb and D. S. Moore (1997) argued that "first courses in statistics should contain essentially no formal probability theory" (p. 820). Since they originally took this stance, the American Statistical Association through their Guidelines for Assessment and Instruction in Statistics Education (GAISE) have endorsed the notion of reducing probability's role in introductory statistics classes (Aliaga et al., 2005). A recent draft of GAISE 2016 lists probability as a topic that can be omitted and that "at most, this course needs basic definitions, the addition and multiplication rules, and the basic concept of conditional probability" (Carver et al., 2016, p. 12). This new call goes against Cobb and Moore's position. Cobb and Moore's position was that there should be little to no emphasis on *formal rules of calculating* probabilities (e.g., what to do for all types of cases for $P[A \cup B]$ or $P[A \cap B]$). They believe that discussing probability is still important for statistical inference. Cobb (2015) reiterated this call by urging statistics educators to strip away formulas so that central ideas can become accessible to students. Liu and Thompson (2002) make an excellent argument that trying to debate the question of "What *is* probability?" is a fruitless endeavor in a first course. Rather, in a first course on statistics and probability, our focus should be on what we (our students and us) shall *mean* by the term "probability".

Students do not learn and instructors do not teach in a vacuum with only each other's company. Course materials such as textbooks play a role in student learning. A quick perusal of four introductory statistics texts indicates that these texts cover probability in the exact way that Cobb and Moore urged against; i.e., each focuses almost entirely on how to calculate rather than how to think about (what do we mean by) probability. I chose these four texts for different reasons. The Samuels (2015) text is the textbook for the course I plan to use as a research site; the Moore (2012) text is one of the most popular introductory texts at the undergraduate level. As we are seeing a surge in number of open access textbooks across all disciplines, I also looked at two such texts (Diez, Barr, & Çetinkaya-Rundel, 2016; Illowsky, Dean, & OpenStax College, 2013). My intent here is to not conduct a full-scale text analysis, but rather just to get a sense of how textbooks treat probability. Worth pointing out is that all four of these texts have publishing dates after the Cobb and Moore article and after the release/adoption of first GAISE report. The introductory text *Statistics for the Life Sciences,* 5th edition (Samuels, 2015) devotes ~20 pages to probability. However, there are only three sentences related to how to think about probability. Out of the 18 exercises provided for the students to use for homework, no question asks students to interpret/make use of a way of thinking about probability; the questions all focus on calculating probability values. Likewise, *Introduction to the Practice of Statistics*, 7th edition (D. S. Moore et al., 2012) devotes 18 pages to probability and randomness. Of these pages, only 3 sentences focus on how to think about probability. There are only two questions of the 45 that focus on something other than a calculation of probabilities or judgment of independence; one asks whether or not a probability value is applicable to a larger set of colleges, and the other asks students to explain what a probability value means. In addition to these two traditional textbooks, I examined two open source texts; *Introductory Statistics* (Illowsky et al., 2013) and *OpenIntro Statistics*, 3rd edition (Diez et al., 2016). There are 51 pages devoted to the topic of probability in *Introductory Statistics*, while the *OpenIntro* gives 40 pages to the "special topic" of probability. There is only one sentence in each that focuses on how to think about probability. For the 128 homework questions in *Introductory Statistics*, only three ask for something other than a calculation of a probability value. Those three questions ask students to state what an expression such as $P[A \text{ OR } B]$ means in words. None of *OpenIntro*'s 44 questions ask students to interpret a probability value. In all four cases, the students' major takeaway is that probability is a calculation.

Chapter 3: Theoretical Perspective

In this chapter I present my theoretical perspective that undergirds this sequence of studies. As a radical constructivist, I believe that individuals build their knowledge based on their experiences. The process of knowledge construction is an active endeavor on the part of the individual and the individual is motivated by a need to maintain equilibrium between their knowledge and their experiences. An individual's knowledge is not a picture of the world; rather knowledge consists of thoughts, images, words, deeds, anticipations, and emotions that organize her experiences as well as being organized by experiences. The primary perspective I use is a theory of meanings. Thompson (2016) presents a theory of meaning based in the traditions of constructivism (Piaget) and radical constructivism (von Glasersfeld). I first present four perspectives on scheme before detailing a theory of meanings. I then propose my particular take on the theory.

Four Perspectives on Schemes

Researchers of cognition often use the notion of "scheme" in their work. However, what these researchers intend to describe and/or convey to readers through their usage of the term "scheme" is not always clear. Here I will focus on four researchers and their usage of the concept of scheme as they looked at students' understandings of mathematical topics. While these four researchers adhere to the (radical) constructivist tradition, there are other schools of thought on schemes. I'll not review these other perspectives but Derry (1996) offers details how the different schools of cognitive science treat the idea of scheme. I will begin by briefly reporting on what each researcher has written as to his meaning of "scheme", followed by highlighting commonalities and differences among them. Finally, I will describe the work that scheme can do for a researcher in understanding, learning, and teaching.

The biologist turned genetic epistemologist Jean Piaget made extensive use of the concept of "scheme" throughout his career. In Montangero and Maurice-Naville's (1997) work, they give multiple quotations for Piaget's extensive works that I find particularly useful. I present three of these quotations below (dates in square brackets are of the original French publications):

> 2. "A scheme is the structure or the organization of actions which is transferred or generalized when this action is repeated in similar or analogous circumstances." (*The Psychology of the Child*, [1966] 1969, p. 11)
>
> 4. "The system [i.e., scheme], composed of determined and completed movements and perceptions, reveals the dual character of being structured (hence of itself structuring the field of perception or comprehension) and of constituting itself from the outset inasmuch as it is a totality." [my addition] (*The Origins of Intelligence in Children*, [1936] 1977, p. 417)
>
> 6. "The scheme of an action is neither perceptible (one perceives a particular action, but not its scheme) nor directly introspectible, and we do not become conscious of its implications except by repeating the action and comparing its successive results." (*Mathematical Epistemology and Psychology*, [1961] 1966, p. 235) (Montangero & Maurice-Naville, 1997, p. 155)

Piaget viewed schemes as aspects of an individual's cognition that served as an unconscious means for the individual to organize and make sense of his/her experiential reality. Schemes for Piaget were not necessarily ready-made recipes an individual followed step-by-step. Rather, schemes served as a general pattern of actions that account for a number of circumstances and could be triggered in multiple ways. For instance, consider the notion of baking. Rather than my wanting to bake always resulting in me making coffee cake, the same set of actions for "baking" can be used to make any number of baked goods such as cheesecake or pie. Additionally, my sense of baking does not necessarily require that a successful (or rarely, unsuccessful) physical product result from my actions; rather, I could imagine what the result might be were I to physically act.

Within the baking example, there is a subtlety highlighted by the third quotation of Piaget (number six). My scheme for baking is not directly observable. While other people might watch me read a recipe, mix together ingredients, put things in the oven, and even consume the end product of my work, but not once can any person say "I see Neil's baking scheme." I cannot even say that I see my baking scheme. Rather, the best anyone could say is "Here is a possible structure that coordinates Neil's actions across multiple instances of baking." Similarly, the best that I can say is that I do certain things because I've compared different experiences I've had when baking.

Item four in the quotation I've provided focuses on the dual nature of schemes and highlights just how central the concept of scheme is to the cognizing individual. We constantly process sensorimotor data; some we focus on, most gets shunted off to the periphery. This sensorimotor data does not remain neurochemical impulses; we make sense of these impulses as we strive to make sense of our environment (physical surroundings and/or mental re-presentations). Yet, in order to make sense of these impulses, there must be something that organizes them. Physically, the brain (central nervous system, more generally) serves this role; however, cognitively, schemes fulfilled this role in Piaget's work. While schemes organize our perceptions, our perceptions, in turn, lead to the evolution of schemes. Through comparing successive results, we may note differences amongst the results and pick up on aspects of the actions that we did not view as being important. These aspects may lead to a change in the current scheme or may even lead to the development of an entirely new scheme. Thus, schemes are not static things, but rather dynamic, evolving things. This aspect should feel as a given as our own experience tell us that how we perceive and make sense of the world around us changes as we grow and have new experiences.

Von Glasersfeld (1995, 2001) drew upon Piaget's works as well as the biological basis Piaget grounded his work in for his written description of the concept of scheme. For von Glasersfeld, a sensorimotor or action scheme consisted of three components as a single cognitive structure. First, the individual must have some recognition of the current sensorimotor data as being that of a certain situation. Second, there must be some specific activity linked with the certain situation that the individual just recognized. Finally, the individual expects that this activity will result in a certain outcome that he/she has previously experienced. Assimilation (another term Piaget used from biology) provides the first component for von Glasersfeld. Von Glasersfeld (1995) described assimilation as the individual viewing new sensorimotor data "as an instance of something known" (p. 62). Assimilation also occurs due to the individual carrying out the second component. The individual attempts to assimilate the actual result of activity to the expectation of the third component of a scheme. If the assimilation is successful, that is, the produced product is consistent with what the individual expects, then the status quo remains. However, if the individual perceives an inconsistency, then an accommodation to the scheme may develop over time. Accommodations include tweaks to the scheme such as attending to new aspects of the original data, revisal of the expected result of activity, and/or the generation of a new scheme. However, such changes do not occur spontaneously or instantly.

Drawing upon the works of both Piaget and von Glasersfeld, Steffe also used the concept of scheme to refer to an organizing structure of the subject's mind. "Scheme" provided him with a way to describe abstract patterns of behavior of an individual across several tasks (Steffe, 1992). Steffe (1983) made use of the idea of an "operative scheme" as a scheme involving mental operations. Here, Steffe used Piaget's concept of mental operations as cognitive primitives. Additionally, Steffe (1983) wrote that mental operations have the attribute of being actions carried out in thought that lead to a result and have content. He uses a counting scheme to further explain the concept of an operative scheme. "Counting as activity has been defined as the coordination of two productive activities. 'Counting is a production of a sequence of number words, such that each number word is accompanied by the production of a unit item (Steffe *et al.*, 1982, p. 83)'" (Steffe, 1983, p. 111). The actual counting activity links nicely with von Glasersfeld's second component of a scheme (i.e., specific activity). What makes this counting scheme operative lies in the second activity. The production of a unit item occurs through the uniting of different aspects into a singular whole in thought; Steffe (1983) refers to this mental operation as "integration". Additionally, the individual has a sense of anticipation of what will be the result should she actually count.

As a student of Steffe, having worked with von Glasersfeld, and extensively read Piaget, Thompson's usage of scheme is an amalgamation the prior three researchers' usage. Thompson (1994, 1996) highlighted the role that imagery plays in a scheme. He noted that Piaget made distinctions between three kinds of imagery. A first kind of image deals with creation of objects (Thompson, 1994, 1996). For instance, imagine a ball; the mental image that you conjured belongs to this kind of imagery. A second form of imagery involves images of the first type with images of actions on the object. Thompson (1996) asserts that in this second form, that "if by actions we include ascription of meaning or significance, then we can speak of images as contributing to the building of understanding" (p. 3). With the ball that you imagined, you can also imagine holding the ball, rotating the ball to see if there is a logo, and tossing the ball. These images are of the second type. Another example of this second type of imagery would involve you ascribing that the imagined ball held a special meaning such as one signed by your favorite athlete. The final type of imagery is that when the individual constructs a dynamic image at a certain moment and that construction is shaped by the mental operations the individual uses. At the same time that these mental operations lead to the construction of the image, the operations are also tested for consistency with the scheme that organizes the operations (Thompson, 1994, 1996). Thus, the development and refinement of schemes occurs over time and with repeated employment. Most recently

Thompson gave a definition of scheme as an attempt to elevate von Glasersfeld's notion to be more encompassing; a scheme is "an organization of actions, operations, images, or [other] schemes—which can have many entry points that trigger action—and anticipations of outcomes of the organization's activity" (Thompson, Carlson, Byerley, & Hatfield, 2014).

A Theory of Meanings

In building a theory of meanings, a number of issue arise. Chief amongst these issues are what the term "meaning" means, how does a person come to have a meaning, and how one individual communicates his meaning for some object to another person. Each of these issues entails many additional ideas and relationships to the others. I'll first present what constitutes a meaning by drawing on the prior section's discussion of schemes. Second, I'll describe how I think that an individual develops a meaning by examining how an individual constructs objects through the perspectives of radical constructivism and symbolic interactionism. Symbolic interactionism and the constructivist paradigm (in particular decentering) play a critical role in how an individual might convey a meaning to another individual. I'll conclude this section by discussing ways to classify meanings in terms of productiveness and usefulness.

A central tenet of radical constructivism is the belief that every individual builds his/her own knowledge through repeated experiences. Through repeatedly reasoning about a set of similar experiences, the individual builds a scheme, which organizes the actions[3] that arose during the repeated experiences. Piaget wrote "A scheme is the

[3] An action is any thought, word, deed, or emotion that fulfills a need, (Piaget & Elkind, 1968).

structure or the organization of actions which is transferred or generalized when this action is repeated in similar or analogous circumstances" (Piaget, [1966-French] 1969 in Montangero & Maurice-Naville, 1997, p. 155). Thus, an individual's schemes enable him/her to make sense of experiences. During an experience, the individual strives to make sense and does so by assimilating the experience to one (or more) of her schemes. When this assimilation occurs, that individual enters a cognitive state in which the individual may now reason about the experience. Thompson and Harel devised a framework that call this state the individual's understanding of the experience (Thompson et al., 2014). Piaget also wrote that schemes are "organized totalities whose internal elements are mutually implied" (Piaget, [1936-French] 1977 in Montangero & Maurice-Naville, 1997, p. 155). Thus, when an individual understands an experience, any actions, images, and schemes associated with the scheme the individual assimilated the experience to can be easily brought to mind by the individual. This is the inference that accompanies assimilation (Jonckheere, Mandelbrot, & Piaget, 1958). I view the inference as a set of implications resulting from the individual's understanding of the experience. In the aforementioned framework by Thompson and Harel, this set of implications is the meaning that the individual gives to experience based upon his/her understanding (Thompson et al., 2014). Given the prior writing about schemes, the meaning the individual has for an experience is the scheme to which the individual assimilated the experience. This view of meaning is also in line with "meanings is a statement of the relation between the characteristics in a sensuous stimulation and the responses which they call out" (Mead, 1910, p. 402). I take Mead's "characteristics" to be the aspects of the situation that an individual assimilates to form an understanding and his "responses" to be the set of associated actions, images, and schemes.

Meanings develop in the same way as schemes; as an individual interacts with his world, he organizes his experiences according to his schemes, and as his schemes develop, he organizes his experiences differently. When experiences fit his existing schemes, those schemes are re-enforced; when experiences do not fit his schemes, he experiences perturbation. To resolve the negative feedback loop that is perturbation, the individual accommodates his schemes, thereby altering the meaning(s) he originally gave to the experience. In this description, the statement "interacts with his world" entails much more than the individual physically experiencing his immediate environment. "His world" consists of objects that he has constructed for himself.

Writing about how a child constructs her reality, Piaget described an object as an individual's construction consisting of "a system of perceptual images endowed with a constant spatial form throughout its sequential displacements and constituting an item which can be isolated in the causal series unfolding in time" (Piaget, 1995, p. 270). Von Glasersfeld (1995) explains that there are two phases to the development of objects: recognizing an object when the individual has sensorimotor data available and re-presentation of the object when sensorimotor data is not available. Mead (1912) viewed physical objects as things a person constructs out of sensorimotor data and his past experiences. Blumer (1986) explained Mead's view of object as "objects are human constructs and not self-existing entities…[objects are] anything that can be designated or referred to" (p. 68). All four of these scholars do not take the objects that comprise an individual's world as a priori givens. Rather they view objects as idiosyncratic to the individual. As an individual builds a conception for a new object, she generates images that eventually allow for the coordination of schemes and the development of her meaning for the object. Von Glasersfeld points out that in order for an individual to construct permanent objects, she must develop a sense of identicalness. In essence, the individual must think of an object as being the same object at different moments. Between the moments, she imbues the object a continual endurance, even when the object is not at the fore of her experience.

When an individual has awareness of what she is doing, thinking, feeling, she has the additional awareness that there is a "she" that is doing, thinking, or feeling. This "she" is her image of self. As von Glasersfeld (1995) noted, the "self" does not need to the product of intense or complicated thought; she will build her model of "self" over time so that her model is viable with her experiences. Mead viewed the self (the "me") as the individual's response to her communication. Keeping in mind that "gesture" includes more than physical motions, Mead (1912) took "any gesture by which the individual can himself be affected as others are affected, and which therefore tends to call out in him a response as it would call it out in another" as a way that the individual builds her "self" (p. 405). Both von Glasersfeld and Mead viewed the "self" as an object that the individual continually revises. While the quotation by Mead conveys sense of what von Glasersfeld referred to as the "social self", Blumer (1986) points out that the "self" is also a mechanism of self-interaction. This encompasses not only the social self, but also the perceived self. Mead's quotation brings in a new issue: the "others". In writing about the social self, von Glasersfeld (1995) said "if it is others from whose reactions I derive some indication as to the properties I can ascribe to myself, and if my knowledge of these others is a result of my own construction, there is an inherent circularity" (p. 127). There is not circularity to this statement, as von Glasersfeld argues, when you examine what constitutes the "others". The "others" are nothing more than objects, models, that we create out of our experiences for people that we believe can behave in predicable ways. Our initial model for the "other" rises out of our model of "self"; however, we adapt the model to allow autonomy in how they react (i.e., the "others" do not necessarily have to respond exactly as the "self" would). As the individual has more experiences, she continues to revise her models of them, always seeking out a viable model. However, as she does this, this also receives feedback as to the viability of her model of herself. This enables her to adapt the "self" so that she can maintain equilibrium between model and experience.

What does the present discussion about objects, self, and others have to do with meaning? When an individual constructs an object, she is creating an element of her experiential world in which she imbues certain meanings. Mead, Piaget, Blumer, and von Glasersfeld would all surely agree that an individual constructs objects through interaction. Blumer (1986) emphasizes that object construction occurs though social interaction (including interactions involving only the "I" and "me"). Defending Piaget's work from the criticism that constructivism ignores the social element, von Glasersfeld (1995) argues that a rich source of perturbations, and therefore opportunities for learning, stems from the individual's experiential world containing other people. Blumer draws upon Mead's work to answer how a person's meaning for any object develops or is refined through social interaction with his notion of symbolic interaction. Symbolic interaction is the portion of social interaction where an individual acts based upon his interpretation of another person's actions. There are two components to symbolic interaction: interpretation, ascribing meaning to another's actions, and definition, the indication of meaning to another person so that she might act. Through social interaction, an individual can enter a state of perturbation, which he would resolve with the modification of existing meanings or the development of new meanings. The individual must make an interpretation of the actions of the other person to give those actions meanings. In other words, the individual must assimilate those actions to his existing schemes. Through this assimilation, he is now positioned to make decisions about his own actions, which are now based upon the meanings he imbued the other person's action with. The individual's actions contain definition of what he wants the other person to understand. However, she must now assimilate his actions to her schemes. However, what happens when she does not act in a way that is consistent with how he imagined? He must now resolve the perturbation that her unexpected (to him) behavior caused with his model of her. Symbolic interaction provides us a start to process of building objects/meanings, but radical constructivism allows us to go further. Interpretation is essentially assimilation. However, there is much more going on with defining. Here the individual must make use his model of the other person with the goal that she will interpret his actions in the way he intended. If he assimilates her actions in a way that supports him in believing that she understood what he intended, then he has evidence to support the viability of his model of her. At the same time, she is doing the same. This state where each participant believes that the other have interpreted his/her actions as he/she intended is what von Glasersfeld referred to as intersubjectivity. There is an important word in this last sentence: believes. Neither individual needs to have a perfect model, nor do they have to have consistent models for intersubjectivity. As long as their models of the other person are viable with their experiences, then they are in a state of intersubjectivity.

Given that an individual develops her meanings through social interaction, the two acts of symbolic interaction serve as means of explaining how a person conveys her meanings to another person. In particular, the act of defining is how she would attempt to communicate her meaning for an object. For the person receiving this communication to have the same meaning as the sender, he must interpret her indication in the way she intended. There is no guarantee that he will interpret her actions in the way she intended and thereby, no guarantee that she has communicated her definition. This brings us to important question in mathematics and statistics education research: If meanings are intensely personal constructions and people involved in symbolic interactions can be in a state of intersubjectivity without having the same meanings, then how can a person learn from someone else? This is a question that Thompson (2013) tackled. Thompson found an answer by turning to the notion of intersubjectivity and Pask's conversation theory. Intersubjectivity hinges on an individual having a mental image of another person that is free to think like and not like the individual (von Glasersfeld, 1995). From Pask's theory, Thompson (2013) highlights that a conversation is more than just a verbal exchanges;

conversation also includes all of the participants' "attempts to convey and discern meaning" (p. 63). He uses the following figure to highlight the blending of intersubjectivity and conversation theory:

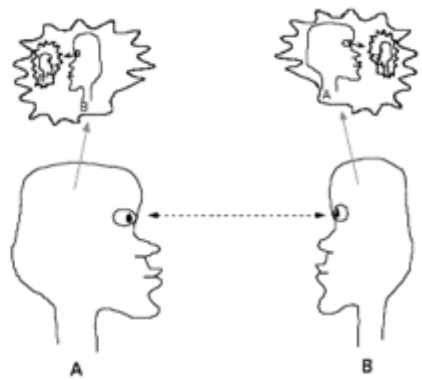


Figure 8. A "meaningful" conversation (Thompson, 2013, p. 64).

As 'A' and 'B' talk to each other, they must each keep in mind not only what they wish to communicate but also how the other person might interpret his/her words/actions. Both 'A' and 'B' build a model of the other person. Thompson's work provides an answer for how the conveyance of meaning from one individual to another might occur. Suppose that 'A' wants to communicate something specific to 'B'. When 'B' assimilates the experience to his schemes, he imbues that experience with a meaning that stems from two sources. The first source is his own meanings; the second is what he knows about 'A'. The meaning that 'B' gives to 'A's utterance or action is what Thompson referred to as the *conveyed meaning.* A conveyed meaning is the set of implications that a receiver attributes to the sender's message constrained by 1) the receiver's de-centering and 2) the receiver's belief that the sender made an honest effort to convey his/her thinking. These constraints lay the groundwork for intersubjectivity and keep both participants in the picture. While 'A's conveyed meaning might not be a perfect reflection of 'A's actual meaning, this is how ‘B’ understood ‘A’ and is the basis for which ‘B’ to now respond. The notion of conveyed meaning is useful in education in several ways. First, we can use this notion in research to attempt to discern what meanings our students have constructed for various topics. Second, we can use this notion in the planning of lessons. By trying to answer the question of “what have I conveyed to my students?” we can engage in de-centering and design meaningful conversations. This second use is easily extended to a third focused on the generation of curriculum materials such as activities and textbooks. With textbooks, we can imagine to types of conveyed meaning; the first being what the authors conveyed to us and the second being what the authors conveyed to our students.

By examining students’ responses, we can characterize those responses by the meaning conveyed. However, to compare categories of conveyed meanings there must be an aspect of the theory that deals with productiveness of the meanings. As Thompson (2016) notes, if researchers seek to design diagnostic assessments to get at the meanings that individuals have, then the researchers must explicate a productive-nonproductive continuum for those meanings and why one meanings is more or less productive than another. I view productive meanings as those meanings that provide coherence to ideas that individuals have as well as affording students a frame to support future learning (Thompson, 2016). Additionally, productive meanings are clear, widely applicable (within reason) across a number of contexts, and rely only on assumptions that the individual can readily express. Alternatively, a nonproductive meaning is a meaning that is almost exclusively tied to the context of the experience and does not support (and perhaps inhibit) future learning. Take for example the following triad of meanings for the sine function:

1. Sine is what you get when you take the Opposite over the Hypotenuse of a right triangle.
2. Sine gives you the vertical displacement from a horizontal diameter of a circle for a given angle measure.
3. Sine describes the co-variation of the proportion/direction of the amplitude (displacement from rest) with respect to the percentage of a period completed (expressed as an angle measure).

The most unproductive of these meanings would be the first one; this meaning hinges on the presence of a triangle and treats sine as the result of what you do with the sides of the triangle. The second meaning is more productive than the first but is still tied to the presence of a circle. While this meaning treats sine more as a function, the students with this meaning can only operate within the context of circular motion. The third meaning is the most productive of the three. This meaning places sine in the broad context of describing periodic co-variation and would enable students to use sine not only in circular contexts (like the second meaning would) but would also allow students to use sine in contexts where there is not a circle such as human physiology (e.g., hip flexion/extension, blood pressure), energy, and tides. Students whose meanings for sine are completely tied to triangles or circles or see the output of sine as being only a vertical displacement will have great trouble reasoning in these contexts. Further, only the third meaning enables students to see that cosine is nothing more than the complement's sine. The meaning for cosine that goes along with the second meaning for sine (i.e., a horizontal displacement from a vertical diameter) creates a schism between the two functions that inhibits students from understanding why you can use either function to develop a model for same periodic co-variation.

Dewey (1910) wrote "Vagueness disguises the unconscious mixing together of different meanings, and facilitates the substitution of one meaning for another, and covers up the failure to have any precise meaning at all" (p. 130). Vagueness serves as an indicator of a nonproductive meaning. I must point out that the present productive-nonproductive continuum focuses on more long-term aspects of meanings. A student whose meaning I would classify as nonproductive along this continuum may be highly "productive" for the student in light of short-term goals such as finishing a homework assignment. *Productive meanings* are clear, widely applicable (within reason), and entail an awareness of and need to explicate any assumptions. As such, productive meanings are not only identified by the researcher but the judgment of "productive/non-productive" comes from the researcher positioning the particular meaning along a continuum of meanings for that concept. I take a *useful-in-the-moment meaning* to be any meaning that allows a student to meet her immediate performance or learning goal. Consider the following meanings for the associative property:

A. Move parentheses.
B. The choice of which of two structures to impose (e.g., $[a+b]+c$ or $a+[b+c]$) does not change the result.

Meaning A is useful-in-the-moment; students can get correct answers. However, this meaning does not necessarily help students when there are more than three terms. However, meaning B is a productive meaning and useful. The label of "useful-in-the-moment meaning" while also applied by the researcher, is much more focused on the here-and-now for the student than that of "productive meanings".

I close with summary of the most important constructs. A meaning is the space of implications (i.e., the scheme) that an individual assimilates an experience to when he generates an understanding of that experience. A conveyed meaning is the meaning that a receiver construes through her understanding of the sender's actions by de-centering and believing that the sender made an honest effort to communicate his meanings. A productive meaning is a meaning that provides coherence to the individual's current ideas as well as serving future learning, is clear (not vague), applicable across multiple similar contexts, and relies on assumptions that the individual can explicate. A useful meaning is any meaning that an individual uses to achieve some goal.

Chapter 4: Conceptual Analysis of Distribution of a Random Variable and a Hypothetical Learning Progression

This chapter contains the conceptual analysis for thinking of the distribution of a random variable as the accumulation of random process outcomes with respect to the value of the random variable. This particular way of thinking about distribution is the target of the hypothesized learning progression that I will use to guide the three studies. As part of the learning progression, I will present progress variables based on upon both the extant research and my conceptual analysis that serve as ways to evaluate students' understandings and place students within the progression.

Conceptual Analysis, The Tool

Before presenting my conceptual analysis, I will first discuss the tool that is conceptual analysis. The phrase "conceptual analysis" refers to an approach for distilling what an individual understands into basic components of the individual's mental actions. While von Glasersfeld (1995) describes his approach, he drew upon the works of Jeremy Bentham and Giambattista Vico as well as his time working with Silvio Ceccato's group. Ceccato's group, "The Italian Operationist School", focused on reducing all words, in any language, not to other words, but rather to the mental operations associated with those words (von Glasersfeld, 1995, p. 6). Von Glasersfeld used the term "concept" to refer to dynamic mental re-presentations that had "been honed by repetition, standardized by interaction, and associated with a specific word" (von Glasersfeld, 1987, p. 219). Through Piaget we can form a link between von Glasersfeld's usage of the term "concept" and Ceccato's group's aim. Montangero and Maurice-Naville (1997) share the following for Piaget's meaning for mental operation:

> 5. An operation as such is a creation of the subject, since it is an action he exerts on things. Action, this being the earliest form of operation, adds new elements to reality…
>
> 6. A system of intellectual operations has two aspects psychologically speaking: externally, it is a coordination of actions (effective or mental actions), whereas internally, that is to say, from the point of view of consciousness, it is a system of relations where each relation implies the others…
>
> 10. Unlike most actions, operations always involve a possibility of exchange, of interpersonal as well as personal coordination, and this cooperative aspect constitutes an indispensable condition for the objectivity, internal coherence (that is, their 'equilibrium'), and universality of these operatory structures…
>
> 11. An operation is not a representation of an act—strictly speaking, it is still an action since it produces new constructions, but it is a "signifying" and not a physical action in that the connections it uses are implicative, not causal (pp. 137-138).

I provided the long quotation of Montangero and Maurice-Naville to convey that mental operations are 1) central to the thinking of an individual, 2) bounded by the individual's experiences, and 3) inter-related. Von Glasersfeld used "concept" to refer to what an individual could mentally re-create for him/herself in the absence of physical stimulus. This fits nicely with Piaget's intent of mental operations, particularly in the sense of coordination of actions and acting as a signifier for the individual. Additionally, von

Glasersfeld's conditions (i.e. honed by repetition, etc.) firmly tie into the exchange aspect of mental operations.

A central premise to conceptual analysis is that the analyst seeks to answer the question "what mental operations must be carried out to see the presented situation in the particular way one is seeing it?" (von Glasersfeld, 1995, p. 78). Given this question and that a researcher may only infer the mental operations of a subject based on the subject's observed behavior, conceptual analysis is a method of modeling building. Thompson (2000) describes three usages of conceptual analysis: 1) to build second-order models for how another person might understand a particular idea, 2) generate a model for meanings that should an individual have these meanings, then that individual is in a beneficial position for future learning and 3) generate a model of meanings that might inhibit the individual in generating an understanding of new situations and/or prevent the construction new, more productive meanings. Thompson (2008) added an additional usage of conceptual analysis; describing the "coherence of various ways of understanding a body of ideas" (p. 45). These four uses of conceptual analysis tie to the goals of conceptual analysis as well. Steffe (1996) identified that conceptual analysis highlights the mathematical reality of a student (that differs from our own) as a valid and authentic construction, identifies the aspects of the student's mathematical reality that function effectively, and suggest what "accommodations of those schemes induced by whatever constraints the student may encounter" (pp. 202-203). The following conceptual analysis of a random variable's distribution involves five additional conceptual analyses dealing with randomness, random variable, random process, accumulation, and probability. As a totality, I (and other researchers) have used all four forms of conceptual analysis in building this conceptual analysis.

Hypothetical Learning Progressions and Progress Variables

Conceptual analyses provide a way to orient a researcher to models of how an individual might think about a particular concept. The conceptual analysis also provides an occasion for the researcher to conjecture as to how the individual comes to have his particular meanings for the concept. Hypothetical learning progressions are models that researchers hypothesize as explaining how a student might learning a particular idea (in a particular way) over time (Duschl, Maeng, & Sezen, 2011). Hypothetical learning progressions are science education's name for the hypothetical learning trajectories of mathematics education. Simon (1995) describes hypothetical learning trajectories as consisting of a learning goal, learning activities, and a hypothesis of how learning will unfold. For learning progressions, there are five elements: identified learning targets, identification of progress variables, the mapping out of progress in stages, operational definitions of learning progress to generate observable learning performances, and assessments to track progress (Corcoran, Mosher, & Rogat, 2009). There are links between the two frameworks. Simon's learning goals are Corcoran et al.'s learning targets. The hypothetical learning trajectory's activities and hypothesis encompass the remaining four elements of learning progression. Common to both trajectories and progressions is that they are empirically testable and contain the natural analogy of following a path. Lehrer (2013) makes a case that learning progressions/trajectories can also be thought of a trading zone. Here, the progression is an opportunity for different constituencies (e.g., mathematics educators, statistics educators, curriculum developers, software design, psychometrics) to share their strengths with each other over the (hopefully) long lifespan of the progression.

Part of a learning progression is the identification of a learning goal. For this sequence of studies, the learning goal will be the understanding of a random variable's distribution as the accumulation of a random process's outcomes organized by the random variable's value. Additionally, the research must identify progress variables. A progress variable represents "(a) the developmental structures underlying a metric for measuring student achievement and growth, (b) a criterion-reference context for diagnosing student needs, and (c) a common basis for interpretation of student responses to assessment tasks" (Kennedy & Wilson, 2007, pp. 3–4). The following conceptual analysis of the learning goal will serve as the basis for the identification of progress variables and stages of progression.

Conceptual Analysis of Distribution

To understand the concept of distribution the individual views the distribution of a random variable as *the accumulation of a random process's outcomes organized by the value of the random variable*. This way of thinking is in line with both von Mises' and Kolmogorov's usages of "distribution". Thinking about a random variable's distribution as the accumulation of random process outcomes with respect variable's value supports individuals in thinking about a function that relates a value of the random variable to a cumulative probability. In other words, this way of thinking about distribution is a way of thinking about cumulative density functions (CDFs).

This target meaning for distribution hints at five concepts that students must develop meanings for and coordinate in order to come to understand distribution in this way. Figure 9 shows these five concepts and the concept of distribution. The gradation of colors supports viewing each of the five concepts as interrelated and that the idea of distribution is the amalgamation of these concepts much like white light is composed of full color spectrum.

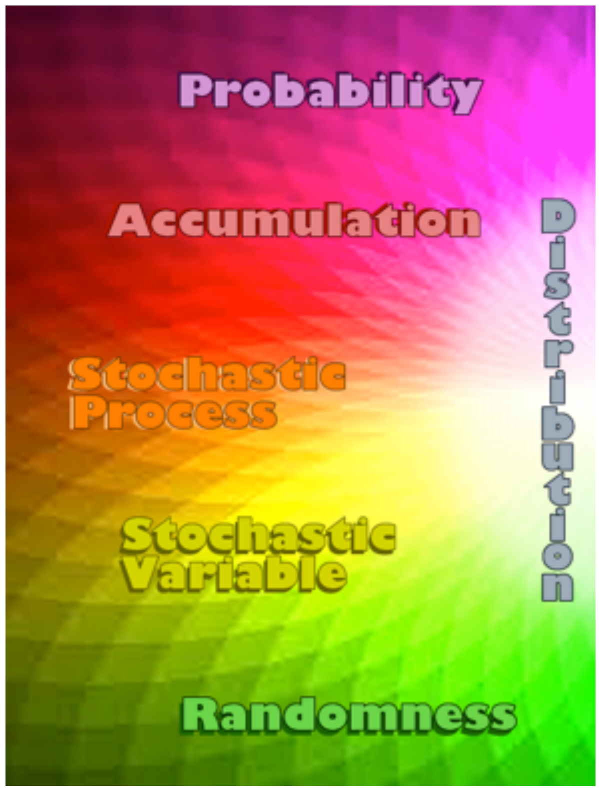


Figure 9. The core concepts for understanding the idea of distribution[4].

The target meaning in this study is to imagine that a random variable's distribution is the accumulation of a random process's outcomes organized by the value of the random variable. Here the student would need to first imagine that there is some process that she can repeat ad nauseam with each trial providing her the value an object's attribute she's interested in. She anticipates that each trial does not always yield the same value as any of the previous trials and that she cannot predict what value any one trial

[4] My thanks to Caren Bergemeister for her assistance creating this figure.

will yield. As she repeats the process, she keeps track of each trial's outcome by building a sequence or list that she adds to with each new trial. While her sequence follows the ordering of her iterations of the process, she imagines that she can re-order the sequence however she wants; she anticipates that the process could have just as easily yield the sequence in any order she imagines. The idea that she could shuffle the sequence into any ordering stems from her imagining that the process is not overtly influenced by factors that she could use to predict the process's outcomes. As she continues to repeat the process and get more values, she begins to anticipate that she can predict how often the process yields different values. By ordering the sequence by the value of the attribute, she can put like values together to make counting easier and so that she can better compare the frequency of individual values. She also imagines that a friend could start carrying out the same process at a different moment and build a sequence in the same way. This would allow her and her friend to compare how often different values occur, relative now to how long they have each been repeating the process. She anticipates that this comparison will allow her to make better predictions for how often the process produces different values. The student encounters an issue in that in order to count or develop a relative frequency, she has to stop or pause the process, otherwise the process continues to generate more values that she puts into her ordered sequence. She imagines that she can use the relative frequencies from an extremely large number of trials to estimate what percent of the infinitely many trials would be for different values. She partners this imagery with the idea that if she starts at the beginning of her ordered sequence (i.e., with the smallest value) and continues through the whole sequence, she'll accumulate an increase percentage of the infinite outcomes. She can continue moving through the sequence, examining larger and larger values, until when she reaches the point where she has accumulated 100% of the outcomes from her infinitely repeated process.

The preceding description unpacks the target meaning and highlights how the five core concepts (random process, random variable, accumulation, randomness, and probability) interplay with one another. In order for a student to think about distribution as described, she would need to have particular meanings for each of these concepts. In the following sections, I'll detail the way of thinking about each concept that I believe is key for understanding distribution as the accumulation of outcomes of a random process organized by the random variable's value.

Random process. One often-overlooked question in introductory statistics courses, is where do data come from? In many of these courses, the teacher provides students with sanitized data collections that nicely exemplify the current topic of instruction and have little to do with real statistical inquiry. However, even in these cases the question of where the data come from is still pertinent for students to wrestle with. The answer to this question is not a data warehouse; rather the answer is the random (stochastic[5]) process.

The student imagines that a stochastic process is a method for obtaining data values that he can use to answer some question about the behavior of a stochastic

[5] I use "random" and "stochastic" interchangeably when discussing processes and variables.

variable(s). In constructing this method, the student needs to first conceive of a stochastic variable that he wants to know something about; for example, the height of people living in the United States who are at least 20 years old. Next, the student imagines the population (20+ years olds living in the USA) and imagines a way of selecting individuals. The student could imagine assigning a unique ID to each person who is at least 20 years old and living in the USA and then placing those IDs into a container. To select an individual the student imagines reaching into container and drawing out an ID; student imagines that he is running a lottery. Once the student has removed the ID, he imagines going the associated individual and measuring (and recording) that individual's height. The student at this point might imagine returning the ID to the container or putting the ID to the side. Whichever option the student chooses, the student anticipates that he could reach into the contain a draw out another ID and measure that associated person's height which the student anticipates as being potentially different from the prior observation(s). The student imagines that this repeatable action is repeatable for an essentially infinite number of trials. That is to say that the student believes that he could carry out this method of collecting heights forever. Each time he carries out the method (a trial), he records the outcome (the observed value of the stochastic variable) as the newest term of the sequence. To describe the above more succinctly, the student thinks about a stochastic process as infinitely repeatable process where each trial yields an observed value of a stochastic variable (the trial's outcome) with the anticipation that the outcomes vary with each trial and the resulting sequence appear random.

There are two types of stochastic processes worth discussing. The first type matches previous example for investigating the height of 20+ year olds living the USA. Stochastic processes of this type help students answer simple questions that revolve around singular stochastic variables and do not necessitate the use of any statistics. I refer to these processes as first-order stochastic processes. First-order processes are sufficient for thinking about the distribution of random variables. However, if the random variable is actually a statistic, then a second-order stochastic process is needed. A second-order stochastic process is a stochastic process that involves carrying out one or more other stochastic processes as a necessary step in carrying out this larger process.

As a first example of conceiving a second-order stochastic process, I turn to Liu and Thompson's (2007) work on teachers' understanding of probability. They provide descriptions of the conceptual operations an individual needs to view a process as [second-order] stochastic. First, the student must imagine that a process exists by which he may find the answer to a question he is dealing with. For example, Liu and Thompson (2007) use the following prompt: "Suppose that 30 people are selected at random and are asked, 'Which do you prefer, Coke or Pepsi?' What is the probability that 18 out of 30 people favor Pepsi over Coca Cola?" (p. 122). While the question deals with finding the value of the probability of 18 out of 30 people favoring Pepsi, the student must first imagine that there is a method by which 30 people may be randomly selected from some population and asked which pop they favor of the two choices. Embedded within this imagined method, the student must also conceive of the existence of a specific population from which 30 people can be asked for their responses. Additionally, the student must anticipate that these 30 people won't necessarily give the same answers as each other nor the same answer the student himself would give. In essence, the student must bring the notion of variation to the foreground in his thinking about the random variable that is cola preference. (This is a first-order stochastic process.) Once he has collected the 30 preferences, the student can find the proportion of people who prefer Pepsi; he finds the value of a statistic that acts as the outcome of the second-order process.

Once the student coordinates a target population, a selection method, and his anticipation of varied responses, he must imagine that he could carry out this same (first-order) process again and again in nearly identically settings; second, given the settings and carrying out the process repeatedly, he must assume that the value of the statistic (the proportion of people who prefer Pepsi) won't necessarily be the same as the prior instances (Liu & Thompson, 2007). This second assumption serves as the basis for a critical anticipation the individual must make. The individual must imagine that just because the first time of randomly selected 30 people from the target population and finding that 14 of the 30 prefer Pepsi, that this does not mean the next time he carries out the process he will also find 14 out of 30 people prefer Pepsi. Additionally, the individual must imagine that since the first instance of process resulted in 14 out of 30 people preferring Pepsi, that this does not preclude any other subsequent instance from also resulting in 14 out of 30 people preferring Pepsi. This anticipation is an extension of thinking about variation between individuals; now the variation is between collections. Rather than just thinking that the selected people won't all give the same answer, the individual must also now imagine that result of each instance of the process (i.e., the proportion of people who prefer Pepsi) may vary much like each person's response.

Just as the student records individual's heights or cola preference to form a sequence, the student can record the outcomes of the second-order stochastic process. The terms of this sequence are the values of the statistic that student used. This record keeping is the final bit of imagery that the individual must incorporate with his view of the two processes (Liu & Thompson, 2007). The individual views the record of results as fulfilling a need so that he may answer the question at hand (e.g., the probability that 18 out of 30 people favor Pepsi). The individual imagines this record as a collection of outcomes stemming from the imagined random process. Acts of generating and adding to this record joined with viewing the record as a collection of outcomes provides the individual with the means to look for patterns that may emerge when carrying out the process a large number of iterations.

Random variable. When the student imagines a random process as the method of by which he gets data, he must already be thinking about what he wants to examine and treat as data. In particular, the student must already have an image of what attribute speaks directly to his research question and have an idea of how to observe/measure that attribute. The student already anticipates that each time he carries out his random process, he'll observe a value of this attribute. The student can begin to think about representing all of these possible values with a random variable. In thinking about the random variable as all possible values for the attribute he's interested in, the student is poised to begin considering the long-run behavior of this attribute.

To begin with, the student must imagine that there is an attribute about an object/living being that he would like to know more about. This starting point is very much in line with what Thompson (1993, 2011) referred to as an individual conceiving of a quantity. The differences between a quantity as defined by Thompson and a random variable come in two areas; 1) the student must view the random variable as not being intricately bound to any one object/being, but rather a class of objects/person, and 2) the student must view the random variable as being inextricably linked to the notation of variation.

First, while multiple objects/living beings can have the same attributes, the individual's image for a particular quantity includes a specific object/person. For the random variable, the student's focus does not necessarily reside on linking the value of the attribute to a specific object/person; rather the attribute's value is of a generalized object. For example, the student might reason that since Sally's height is 64 inches, she is two inches taller than Pamela. Now, suppose the student thinks of a height of some person that is 64 inches. The student's focus is no longer coordinating a specific person with that height. Rather, this height (64 inches) belongs to an amorphous person who might not be Sally; the student's mental image could be of any other person, man or woman, who he/she feels fulfills the 64 inches tall requirement.

Second, the individual must imagine that random variables are inherently linked with the notation of variation. Students must view variation as changes in the value of some attribute. However, the way students think about "changes" is different for random variables than quantitative variables. For a quantitative variable, the student might think of a change in the value being the difference between the attribute's measure at two different point in time. For example, the height of a rock tossed into the air after 0.5 seconds from release and 2.5 seconds after release. For a random variable, the student needs to also think about changes in the value of the attribute as being the result not of the passage of time but the student looking at different object/living being. For example, the student might see that a person has an A+ blood type and an age of 32 years. The student then looks at different person and sees that this person has a blood type of B+ and an age of 29 years. The student must realize the "changes" in the values of blood type and age (since birth) did not result from the passage of time as in the case of the rock. This form of variation is what I refer to as "variation between individual objects/beings in a collection". I refer to the variation in the tossed rock situation as "variation within an individual object/being". These two phrases help to highlight the major distinction between quantitative variables and random variables; random variables deal with many objects and living beings while quantitative variables deal with a single object or being. The student's anticipation of variation in the values of the random variable feed back into the notion of randomness. The variation prevents prediction in the short-term but not in the long-run.

So far, I've not made a distinction between stochastic variables that are categorical or numeric in nature. I do not use the term "value" to refer to solely numeric measures. The above way of thinking about random variable works for both numeric and non-numeric attributes. The student need only anticipate what values he/she could observe for the random variable of interest to deal successfully negotiate the differences between the two cases. In a similar fashion, there is a distinction between discrete and continuous random variables. If the student conceives of the random variable in such a way that he imagines that the numeric values tell him the count of something (e.g. the number of viable eggs, number of people who prefer Pepsi to Coke) and that there are gaps between successive (adjacent) values, then he is imagining a discrete random variable. However, if the student imagines that the values tell him the amount of something (e.g. a person's height or weight) as well as imagining that values lie along a continuum that he anticipates constantly zooming in upon to find more values between what appeared to be adjacent values, then the individual has conceived of a continuous random variable. This anticipation of possible values brings the student's understanding of the attribute and situation to mind, enabling him to think about what values are possible and make sense. An adult elephant weighing in at 3 kg makes as much sense as a blood type of polka or π. In all three cases (non-numeric, discrete, and continuous), the student must coordinate an imagined generic object/being, variation, and admissible values to truly conceive a random variable.

Accumulation. Stochastic processes provide the way of getting values of a stochastic variable and terms of our sequences. Act of recording these values is an act of accumulating outcomes of the stochastic process. Thompson and Silverman (2008) pointed out that the idea of accumulation is both easy and difficult for students to conceptualize. While they focused on the conceptions of accumulation in the context of calculus, their work serves just as well in context of the distribution of a random variable. In fact, both the trivial and difficult aspects of accumulation as laid out by Thompson and

Silverman (2008) are necessary ways of thinking for the individual to develop and coordinate.

Thompson and Silverman (2008) wrote that trivial conception centered on the image that a student "accumulate[s] a quantity by getting more of it" (p. 1). This image is rather intuitive for students. This conception of accumulation participates in the random process scheme; particularly in the construction of a collection of outcomes of the process. As the student imagines repeatedly carrying out the process and keeping track of the results, he essentially imagines the outcomes accumulating. Based on my experiences teaching, students tend to use a list to keep track of the accumulating outcomes for a random process. With each new trial, the student adds a newly observed value of the random variable to his list. This dynamic image of accumulation of outcomes of random process changes for the student as he envisions the results of new iterations of the random process. The student anticipates that the value he imagines adding to the list each time could be smaller, larger, or the same as the previously added value. Thus, the student imagines the accumulation unfolding in a haphazard way. Depending on the nature of the instruction and the availability of tools, students can be prompted to convert their list into any of a number of data visualizations such as dot plots and histograms. Even in these visualizations, the students can still see the haphazard accumulation of outcomes. Figure 10 provides a series of three images that correspond with haphazard accumulation.

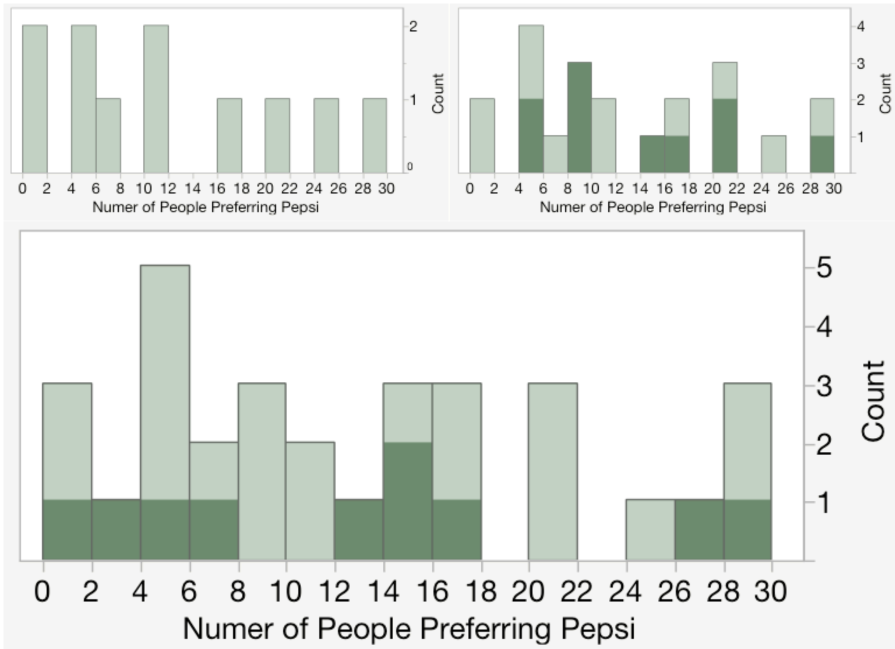


Figure 10. Haphazard image of accumulation of outcomes of a random process. Dark green denotes the 10 new instances of the process added to the previous collection.

The upper left panel shows an initial set of eleven outcomes for the number of people (out of 30) who prefer Pepsi to Coke. The upper right panel gives an image of the collection when the student has imagined running the second-order random process another ten times. The dark green coloring denotes the ten newly observed outcomes. Notice that the dark green bars appear in multiple locations; there were two new outcomes for the values of 4-5 people, three for 8-10 people, and 20-22 people and one new outcome for each of the intervals 14-16, 16-18, and 28-30 people. The final panel shows an additional 10 trials of the second-order random process. The dark green highlights these new ten outcomes (the prior new cases are now part of the light green). Notice that the location of the new outcomes is not the same as in the upper right panel; the accumulation of outcomes followed the flow of stochastic processes. While this intuitive view of accumulation has a place in conceiving the distribution of a random variable, a second conceptualization of accumulation is also needed.

Thompson and Silverman (2008) note that the difficult side of accumulation occurs when students "cannot conceptualize the 'bits' that accumulate" (p. 1). They go on to write that when students coordinate the values of x, f, and total bounded area from an initial value of x to the current value of x, then the student has conceived of an accumulation function (Thompson & Silverman, 2008). This coordination is dynamic for the student; as the student imagines the value of x changing through the domain, the value of f changes accordingly, and the value of the total accumulated quantity changes simultaneously. The imagery involved here in terms of the distribution of random variable requires the student to make several jumps. First, the individual must imagine the repeatedly carrying out of the random process a large number of times *as having already happened.* Returning to the Pepsi vs. Coke example, the student must imagine a large collection of outcomes as existing even if the individual has not actually constructed/simulated a large collection. Second, the student must coordinate the value of the random variable with the accumulated collection of outcomes. In the haphazard accumulation image, the value of the random variable functions more as "binning" classifier; that is, the way to construct the bars for bar charts (non-numeric values) or histograms (numeric values). Once the student places an outcome in the proper bin, the value of the random variable does not serve much more purpose. In this new image, the outcomes already exist, and the accumulation of outcomes occurs as the individual imagines running through the values of the random variable sequentially (with respect to the variable and not the order of trials). Figure 11, while an image a cumulative density curve, serves as good image of the second form of accumulation a student needs to conceive of for the distribution of a random variable. Connecting back to Thompson and Silverman (2008), the domain of the random variable consists of the student's anticipation of

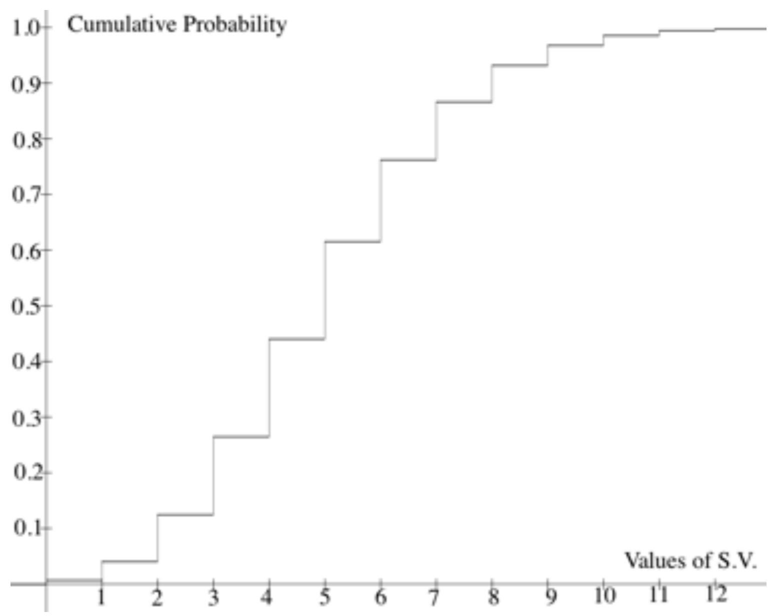


Figure 11. Accumulation of outcomes of a random process with respect to the value of the random variable.

admissible values for the random variable. The student imagines moving through these values in an ordered way. For non-numeric values the student will need to impose some order; for numeric values, the student can use the natural ordering of smallest to largest. As the student moves through the domain of the random variable, the student can now imagine tracking the accumulation of outcomes that are at most as large as the current value of the random variable. Using Figure 11, when the student moves to the value of 2, the student still has the all of the outcomes that he accumulated when he considered the value of 1. This image of accumulation is much more structured than the haphazard view and ties strongly to the stochastic variable. The student begins to form the anticipation that after a certain value of the random variable, he will have accumulated all of the outcomes.

Randomness. At this point in time, the student wants to answer some question about the long-run behavior a random variable and imagines using a random process to get accumulate a list of observed values. To have confidence that his trials will actually allow him to answer his question, the student must have confidence that his list allows for long-run predictions and does not misrepresent the actual behavior of the random variable. This is where the concept of randomness comes into play. In order to think about distributions in a way consistent with the target meaning, students need to have a meaning for randomness that supports the students in viewing randomness as something that they can control and as necessary. If students think of randomness as being "unpredictable", then anything built on this idea automatically inherits this lack of control. A useful way to think about randomness is for the student to think of randomness as an attribute of their list of that 1) minimizes sources of bias and 2) enables the student to make claims about the long-run behavior of the random variable. The minimization of bias supports the student in view the random process as not being influenced by outside forces. In this way, the student views randomness as a way to strip away causal/deterministic effects that would control the variation in the random variable. The second aspect of this way of thinking about randomness runs counter to common meaning for randomness as being unpredictable. While the student acknowledges that short-term prediction, such as the outcome of the next trial (i.e., the next term in the list) are impossible to do with complete accuracy, he anticipates that he can predict the how often the random variable takes on certain values over a large number of trials. Thinking about randomness as a way to retain some control over random variable, that is, they can predict what percent of the time certain values occur, supports students in having confidence that their list can actually help them answer their research question.

In viewing randomness as a desirable attribute of a list of values built from carrying out a random process, students must wrestle with how to check for whether or not a list has this attribute. Students can check a sequence for randomness by examining sequence in three ways. First, the student may look for a pattern that describes the sequence. This approach is natural one for students. If the student can come up with a pattern, then the student declares that the sequence is not appear random. The patterns that students look for in this method are based on the position of each term in the list. However, there are often much more complex patterns that are not detected by just looking at sequence of values in the list. This is where the methods generated by Kolmogorov (sequence complexity) and von Mises (principle of the impossibility of a gambling system) come into play. These other two methods require appear to require significant instruction for students to incorporate into their thinking about randomness. The second way a student can check a list hinges on Kolmogorov's notion of complexity and is in line with notions that Falk and Konold (1994, 1997) put forth. Here the student would need to come up with the shortest description of the list that another person could use to perfectly recreate the list. The more complex that description is, the more the student should accept that the list appears to be random. Interesting to note is that when first introducing the notion of list complexity in the Fall 2016, a group of students came up with idea to time how long it would take someone to come up with his/her description; the more time the person needed to write the description, then the "more random" the students judged the list to be. Descriptions that are insufficiently complex often point to patterns that drive the sequence. The third way to check for randomness stems from von Mises'(1981) Principle of the Impossibility of a Gambling System. If the student can impose a way of picking/discarding terms in a list of fixed length leads to changes in the relative frequencies of values, then that student states that the sequence is not random.

Probability. Imagining the accumulation of something brings to mind notions of how to measure that accumulation. There are four ways (Figure 12) that a student could imagine measuring the accumulation of the outcomes of a random process. Each of these ways requires the student imagine the accumulation as having already happened and ordered by the value of the random variable. In terms of the accumulation of random process outcomes, the most natural way to measure the accumulation is to simply count the number of outcomes for each value of the

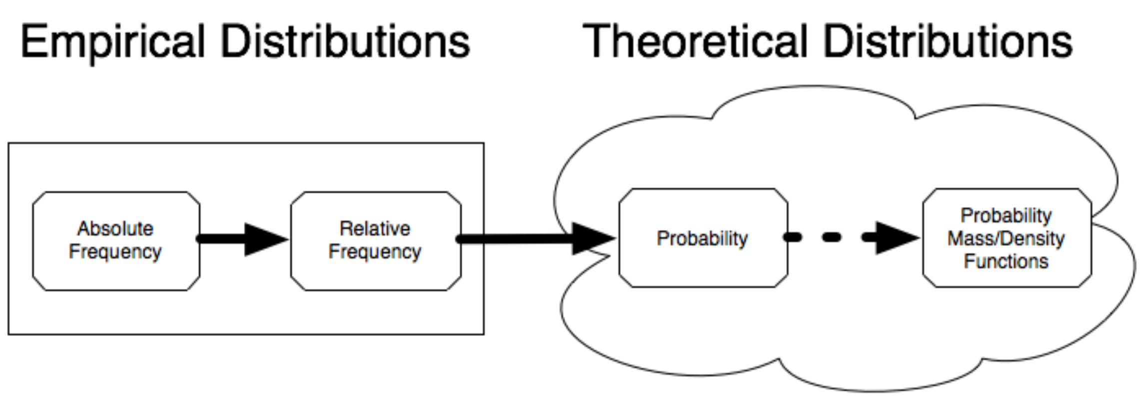


Figure 12. Four ways to measure the accumulation of random process outcomes.

random variable. This is the absolute frequency box of Figure 12. However, if the student wishes to make comparisons with another students, then absolute frequency is not the best choice, especially if the two students did not conduct the same number of trials. Adjusting for the number of trials leads the student to measure the accumulation with relative frequencies. Both the absolute and relative frequencies require a fixed number of trials of the stochastic process; this supports the students in constructing an empirical distribution of the random variable.

If the student imagines the stochastic process running indefinitely, he can then form an anticipation of the percent of the time that stochastic process produces a particular outcome. This percentage of the time or the long-run relative frequency the student thinks about the probability of an outcome occurring, given his assumptions about the stochastic process. When the student makes the jump to using probability to measure the accumulation of outcomes, the student has moved into the realm of theoretical distributions of random variables. Usage of probability is one that is cumulative; thus, the student begins to develop an image for distribution that is consistent with the notion of a cumulative density function.

The fourth way of measuring accumulation moves beyond the cumulative density function notion and towards the ideas of rate of change, as denoted by the dashed arrow in Figure 12. Thompson and Silverman (2008) point out that whenever "something accumulates, it accumulates at some rate" (p. 11). Accumulation and rate of change are different sides of the same coin. Thus, if a student conceives of the distribution of a random variable as the accumulation of outcomes of a random process with respect to the value of the random variable, then the student should also be able to conceive of the rate of change of accumulation of outcomes of a random process with respect to the value of the random variable. Probability mass and probability density can serve this role. In conceiving probability mass/density as the rate of change of accumulation of outcomes of a random process with respect to the value of the random variable, the student must "coordinate images of respective accumulations of accruals in relation to total accumulations" (Thompson, 1994, p. 6). Probability mass and probability density functions describe how the accumulation of outcomes increases as the value of the stochastic variable increases. The value of the random variable accrues leading to a total accumulation of values of the random variable in tandem the accumulation of outcomes of the random process linked with accruals of outcomes. When the student deals with a continuous random variable, he needs to think about the changes in the random variable value as happening in little bits (as he would need to in calculus) so that the accumulation of probability is essentially linear with respect to the random variable. The rate of change for these small intervals is a value of probability density. The student anticipates that the accumulation happens in a smooth, continuous way. In the discrete case, the student acknowledges that smallest change in the value of the stochastic variable is fixed to the size of one step. There are no values between two adjacent values. Thus, the student anticipates the accumulation of outcomes in halting chunks. The rate of change here is probability mass at each particular value of the random variable.

Progress Variables and the Hypothesized Learning Progression

From the conceptual analysis, I propose the following progress variables, presented as construct maps (Wilson, 2005, 2009, 2012): randomness, random variable (including variation), random process, accumulation, and probability. The levels within each construct maps form the base components of the proposed learning progression. For each of these construct maps, the most desirable way of thinking about the concept appears at the top of the table with each following level being viewed as a less productive meaning.

Random process. The construct map for random process (Table 2) has at the very bottom the case where a student rejects the notion that the behavior of random variables and statistics can be modeled. More generally, this way of thinking would support the rejection of building any models in any context. While located directly above this way of thinking is the Deterministic Model meaning, I cannot stress enough that the distance between these two ways of thinking in terms of productivity is vast; one hundred blank pages between these two levels would still not be enough to convey the differences in productivity. In terms of the development of the concept of distribution, the meanings of first- and second-order models are sufficient for thinking of distribution as the accumulation of random process outcomes. To a certain extent, the chance model could also be sufficient. However, the chance model runs the risk of supporting the student in clinging to the "Principle of Ignorance" (Weisberg, 2014) in assuming that every outcomes is equal chances of occurring. Additionally, if the student does think about random processes as chance models, the student feels no need to imagine carrying out the process any number of times, let alone imagining the process running indefinitely. However, chance models are not deterministic models. While the student does not feel the need to carry out the random process, he envisions for the chance model, he anticipates that he would observe variation in the outcomes just as he would if he were to think about a first-order random model.

Table 2. Random Process Construct Map

Second-Order Random Models The student imagines a method of generating values of a statistic of interest to answer a research question about a population. In generating the values of the statistics, the student envisions a first-order random model to get data values necessary for the generation of the value the statistic. The student anticipates being able to repeat this larger process of generating values of the statistic indefinitely and that the values will not always be the same. Adapted from (Liu & Thompson, 2002).
First-Order Random Model The student imagines a method of generating values of a random variable of interest to answer a question about that random variable in some population. The student's image includes the carrying out the method infinitely many times and expecting variation in outcomes of each trial of the method. The outcomes are values of the random variable and form the sequence that the student checks for randomness. Adapted from (von Mises, 1981).
Chance Model The student imagines a method of generating values of the random variable, while the student could carry out an infinite number of times, the student does not feel the need to carry out any trials of the method in order to answer questions about the random variable. The student is able to completely specify each and every outcome without running trials and the student assumes that each outcome has the same chance of occurring as every other outcome. Adapted from (von Mises, 1981; Weisberg, 2014).
Deterministic Model The student imagines a process where he/she anticipates what the result will be before carrying out the process. Further, the student anticipates that if he/she carries out the process under identical conditions again and again, the result will be essentially the same each time.
Null Model The student believes that there is no way to model the behavior of a random variable or statistic.

Random variable. Table 3 shows the construct map for random variables and variation. Rather than forcing a separation of the notion of variable and the notion of variation, I find keeping the two ideas together more useful. The construct map in Table 3 is applicable to variables in multiple mathematical contexts such as (school) algebra and calculus. As I mentioned earlier, the major differences between two is that the stochastic variable is not tied to any one particular object and the dominant image of variation is between individual objects/beings rather than within an individual object.

The first difference is a result of the second different; in order to imagine variation between objects, the student must think of the random variable as connected to many different objects at the same time. While Between Object variation is the primary image of variation, this does not mean that we do not want students to exclude Within Object variation from random variables. In fact, the coordination of the two is critical for longitudinal studies. The mathematical (quantitative) variable the student will envision the systematic (Within Object) variation and a deterministic process for thinking as the mathematical variable as a function. The bold headings of Table 3 are the levels of the construct map; the italicized headings indicate an image of variation that can occur at that level, except where otherwise indicated. For the target meaning of distribution of a random variable, student need only think of random variables at the Variables Vary level with an image of Between Object Variation. However, however thinking of the random variable as a function of a random process can help strengthen the student's understanding of distribution. Thinking of a random variable as a function is necessary for the student to conceive of the distribution of a statistic; the statistic (a function of data) now plays the role that the random variable did. The value of the statistic is the result of a second-order random process.

Accumulation. The accumulation construct map appears in Table 4. While both meanings are useful for the student to develop the idea of distribution, the haphazard image of accumulation is insufficient for viewing distribution as the accumulation of random process outcomes with respect to the value of the random variable. The students need to make the jump from in-progress accumulation to completed accumulation.

Table 3. (Random) Variable Construct Map

<table>
<tr><td colspan="2">Variables as Functions
The student views the variable as result of applying a function to the value of another variable. The student envisions a set(s) of possible values that the variable(s) may equal.</td></tr>
<tr><td>Deterministic Process
The student anticipates that the function uniquely determines the value of the variable given the value of the second variable.</td><td>Random Process
The student anticipates that the function is the application of random processes.</td></tr>
<tr><td colspan="2">Longitudinal Variation (Special Case of Variables as Functions, not a subtype)
The student views the variable as a composition of functions such that there is a coordination of Between Object Variation and Within Object Variation.</td></tr>
<tr><td colspan="2">Variables Vary
The student views the variable in the Measure sense as well as anticipating that the value of the variable changes.</td></tr>
<tr><td colspan="2">Between Object Variation (Element-wise Variation)
The student envisions that the variable’s value may/will change as the individual shifts his/her attention from element-to-element of a set of similar objects all possessing the same attribute.</td></tr>
<tr><td>Within Object Variation—Systematic
The student envisions that the variable’s value changes with uniform motion such that the individual views each value as representing the measure of the quantity of the same object at different moments in time.</td><td>Within Object Variation—Haphazard
The individual envisions that the variable’s value changes at whim through the jumps/drops in values.</td></tr>
<tr><td colspan="2">Measure
The student views the variable as representing the measure of some object’s attribute.</td></tr>
<tr><td colspan="2">Fixed Number
The student views a variable as some fixed known/unknown number. If known that number is known, this leads the student to plug-and-chug calculations; if the number is unknown, this student attempts to solve for the value that satisfies presented conditions.</td></tr>
</table>

Table 4. Accumulation Construct Map

<table>
<tr><td>Finished (Ordered/Systematic) Accumulation
The student imagines the infinite accumulation as already have happened and imagines describing the accumulation of outcomes by moving through the all possible values of the stochastic variable in an ordered and systematic way.</td></tr>
<tr><td>Haphazard Accumulation
The student imagines the accumulation of outcomes unfolding as he/she imagines carrying out the random process. The values are added to a list that is organized by trial rather than by value of the stochastic variable.</td></tr>
<tr><td>Other
The student’s image of accumulation does not fit the other levels.</td></tr>
</table>

Randomness. The construct map for randomness appears in Table 5. The most desirable way of thinking about randomness is that of viewing randomness as an attribute of a process. This way of thinking extends the sequence complexity conception through the inclusion of the usages of randomness.

Table 5. Construct Map for Randomness.

<table>
<tr><td colspan="2">Attribute
The student thinks of “randomness” as a property of a list/sequence/process that entails an image of unpredictability in short-run, while anticipating the predictability in the long-run and minimizes sources of bias. A random sequence has no discernable pattern, has a sufficiently complex description, and adheres to the Principle of the Impossibility of a Gambling System. Adapted from (Kolmogorov, 2013; Liu & Thompson, 2002; von Mises, 1981).</td></tr>
<tr><td colspan="2">Sequence Complexity
The student thinks that a list/sequence as being “random” if the individual’s attempt to describe the list/sequence is to essentially repeat the sequence as given. The individual cannot condense/reduce the list/sequence to a pattern or set of rules that is less complex than the sequence as given. Drawn from (Falk & Konold, 1994).</td></tr>
<tr><td colspan="2">Left-field
The student thinks that events such as sudden switches in conversation topic, unanticipated question, and unexpected images as being “random”. Inspired by (Liu & Thompson, 2002).</td></tr>
<tr><td colspan="2">Unknown/Unpredictable
The student thinks that a “random” event is equivalent to not knowing or being unable to predict the result. For example, upon hearing knocking on a closed door, a student with this way of thinking will say that some “random” person is at the door since he does not who is at the door. Drawn from (Saldanha & Thompson, 2014).</td></tr>
<tr><td>Ordained
The student thinks that “random” events are the result of a chain of events that are meant to occur. Thus, the student believes that nothing is random</td><td>Chaos
The student thinks that all events are random and that whatever happens is the result of happenstance.</td></tr>
</table>

Probability. Table 6 shows the construct map for probability and has appeared in other work (e.g., see N. J. Hatfield, 2016a, 2016b). For the present meaning for distribution, a student needs to think about probability as the long-run relative frequency at minimum. If the student makes the jump to probability mass/density, then student still has the necessary meanings to construct the target meaning for distribution. The occurrence of the Circular meaning grew out of prior research where ~78% of students (and two university instructors) made statements that conveyed the circularity of terms after the students had already received instruction on probability (N. J. Hatfield, 2016b). While additional research is needed to verify the presence and scope of Circular meanings; this result is not surprising given that everyday language treats the terms “probability”, “likelihood”, “chance”, and “odds” as synonyms and some introductory texts explicitly do the same. However, statisticians do not use these three terms interchangeably (and especially not “odds”). “Likelihood” inverts the relationship between data and assumptions that is present in probability. Probability is the long-run relative frequency of data given assumptions, while likelihood is the long-run relative frequency of our assumptions being true, given the data we have on hand. The distinction between probability and chance is one of value generation. Chance grows out of chance models and adheres to the Principle of Ignorance. Von Mises (1981) spends several pages highlighting the distinction between probability values generated via infinite runs of a random process and values generated under set theory axioms. Rather than using “probability” for both set theory based values and long-run relative frequencies, I argue that using two different words (“chance” for set theory and “probability” for long-run relative frequency) is better for helping students develop productive meanings for both ideas (N. J. Hatfield, 2016a).

Table 6. Probability Construct Map

Probability Mass/Density The student views a probability mass value as both the long-run relative frequency and an amount of accumulation with respect to the value of the random variable. The student views a probability density value as the rate of change of outcome accumulation with respect to the value of the random variable.
Long-run Relative Frequency The student views a probability value as the percent of the time he/she expects to see some event when carrying out a random process an infinite number of times. This value is the long-run relative frequency of that event given the student's assumptions about the situation (including the random process). Adapted from (Kolmogorov, 2013; von Mises, 1981)
Relative Frequency The student views a probability value as being the number of times an event occurs relative the number of times the associated random processes has run.
Classical (Laplacian/Chance) The student views a probability value (particularly when displayed as a fraction) as consisting of two number, the first number telling how many ways there are to observe the event out of a total (fixed) number ways to see all outcomes (the second number). Adapted from (von Mises, 1981; Weisberg, 2014).
Prediction The student views a probability value as an assessment of whether the event will occur on the next iteration of the random process (Konold, 1989) and/or the value is a long-run prediction of an unspecified thing.
Fixed Observations The student views the probability value (particularly when displayed as a fraction) as consisting of two numbers. The first number represents the exact frequency of the event while the second number is the fixed number of times the random process has run. The student anticipates that the random process could be run more times, but in sets the size of the denominator. For example, a student would interpret $\frac{3}{36}$ as "every time we roll two dice 36 times, we will get a product of 4 exactly 3 times. If we roll the dice another 36 times, we'll see 3 more products of 4."
Circular The student views probability as a "measure of likelihood/chance" where "likelihood" and "chance" lead to each other and/or back to "probability". The student essentially views probability as a set of word substitutions that can include "chance", "likelihood", "probability", and "odds" among other words and phrases.
Other The student's understanding of probability does not fit any of the other levels.

Hypothesize learning progression. Drawing upon the levels of the concept maps for the five progress variables leads to my hypothesized learning progression for the concept of distribution as shown in Figure 13. I've represented each progress variable (and distribution) with a different shape (mid-right). In the upper-right, the lowest levels of progress variables appear with few hypothesized links between these meanings. Students with these meanings will have little success in building a meaning for distribution that entails the accumulation of random process outcomes with respect to the value of the random variable. On the left side of Figure 13, we see many more levels of the progress variables and the hypothesized links between those meanings. Meanings with a single star are those meanings that I've identified through the conceptual analysis of distribution as being necessary for a student to come to think of distribution in the way described. Meanings with the double star are meanings that go above and beyond the necessary meanings. These meanings would support the students in construction a meaning for the sampling distribution of a statistic. The solid arrows indicate that I hypothesize that meaning at the start of the arrow supports the student in constructing the meaning at end of the arrow for another concept. The dashed arrow shows a similar hypothesized connection but between meanings for the same concept. For example, I hypothesize that a student needs to have a meaning of relative frequency before he/she can development the long-run relative frequency meaning for probability. The dashed line with semi-circle ends indicates a hypothesize impediment for the student developing a more productive meaning. I hypothesize that student whose meaning for probability is in the Classical ("Chance") category will struggle to come to think about probability as a single value (rather than two numbers joined by a bar) that represents a long-run relative frequency. Through this dissertation I aim to refine the levels of the progress variables and the hypothesized learning progression. To go along with the learning progression

(Figure 13), I propose an initial progress variable/construct map for the concept of distribution in

Table 7. The top two levels of this construct map coincide with the two "clouds" that appear in Figure 13. The lower four levels reflect the usages of "distribution" in the extant statistics education literature (see Chapter 2). This construct map is a starting place for how people might think about distribution. I believe that what instruction a student receives has a direct impact on what meaning she constructs for distribution. Several of the levels I suspect are purely byproducts of instruction and do not reflect meanings that students would build through their experiences in any other setting (e.g., the Features level). My intent here is to give myself a sacrificial progress variable that I can improve through this dissertation study.

Table 7. Distribution Construct Map

Sampling Distributions The student views a distribution as being the accumulation of a second-order stochastic process's outcomes with respect to the value of a statistic. The student sees that the values used in calculating each value of the statistic as being the outcomes of a first-order stochastic process for a random variable.
Distribution The student views a random variable's distribution as the accumulation of outcomes from a first-order stochastic process with respect to the value of the random variable.
Collection The student views a distribution as being a collection of values with no/limited image of an underlying process.
Shape The student views a distribution as the shape that they see in a data visualization.
Arrangement The student views a distribution as a particular arrangement of data values along a number line.
Features The student views a distribution as consisting of only measures of central tendency, spread, and shape.
Other The student's understanding of distribution does not fit any of the other levels.

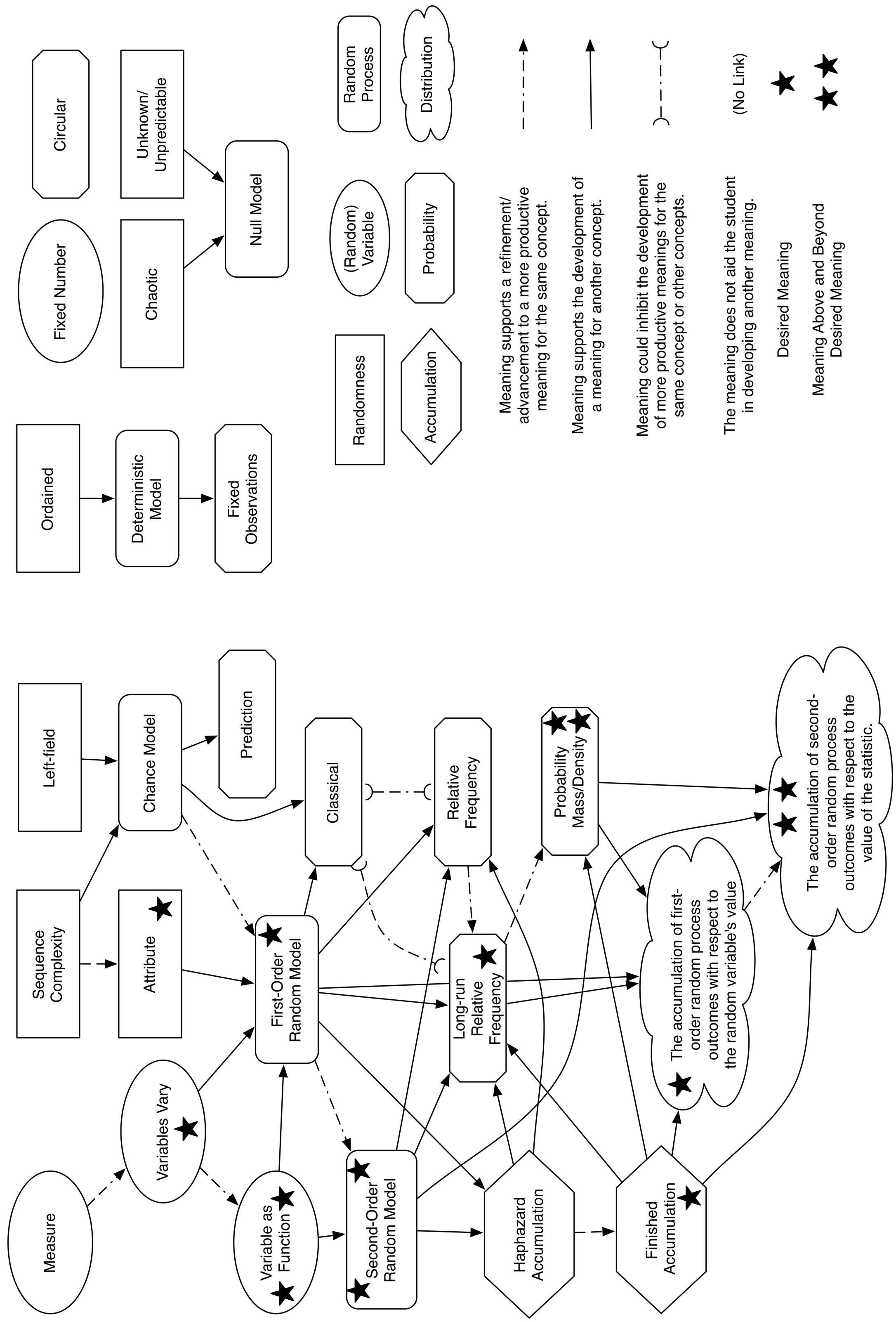


Figure 13. Hypothesize learning progression for distribution.

Coherence and Fit

The proposed meaning for distribution provides a level of coherence that other treatments of distribution do not. The proposed meaning for distribution brings several key ideas together in a unifying way. Should a student develop this way of thinking about distribution, then the notions of randomness, random variable, random process, accumulation, and probability come together as cohesive whole. Treating distribution as an arrangement of values leaves randomness, random process, and accumulation out in the cold. The proposed meaning for distribution is consist with both von Mises's and Kolmogorov's usages of the term "distribution" and supports students in constructing cumulative density functions as an almost immediate extension. By having students develop the notion of CDF out of the target meaning for distribution, the students stand a better chance for engaging in emergent shape thinking rather than static shape thinking (K. C. Moore & Thompson, 2015). By leaving the random process out of the development of distribution, and focusing solely on static graphs, students run the risk of confusing the graph of the CDF (and PDF) with the actual distribution. This could lead students to produce similar statements for the following two graphs:

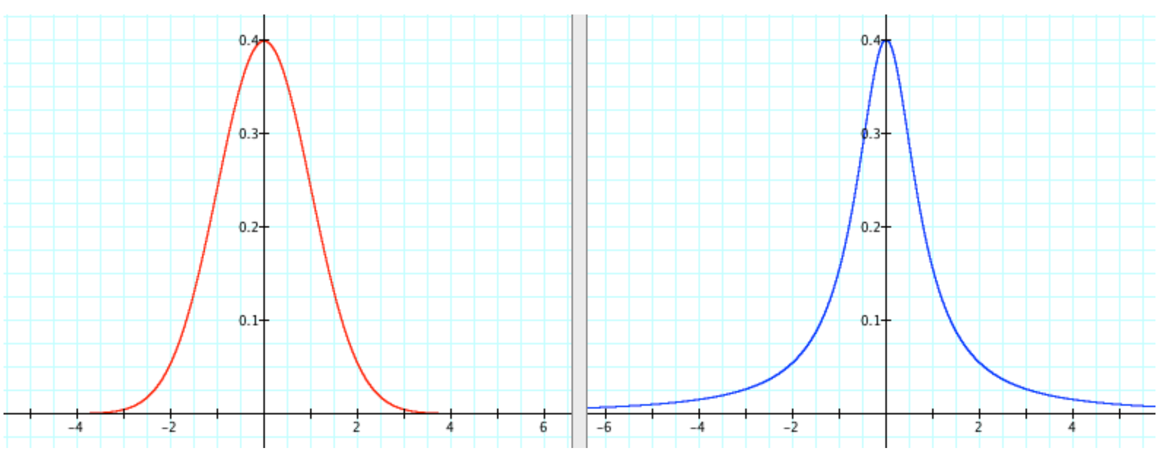


Figure 14. Two graphs of probability density functions.

The student might claim that in both cases, distribution has an expected value (mean) of zero. However, this is a category error. The graph on the left is the graph of a probability density function for a random variable that has a Standard Normal distribution. The graph on the right is of the probability density function for a random variable with a Cauchy distribution; a distribution that has an undefined expected value. Providing students with opportunities to learn about increasingly sophisticated stochastic processes and supporting the students in keeping those processes in mind when developing distributions can help students keep from making such mistakes. My target meaning for distribution could, as part of a larger sequence, help students develop an arsenal reference situations that Tukey (1975) viewed as being a key starting point for statistical investigations. Further, this way of thinking about distribution lays a natural extension via second-order stochastic processes into dealing with sampling distributions and statistical inference.

Similar to the aforementioned, the target meaning for distribution affords students the opportunity to 1) conceive of a new object that has attributes, and 2) see where *p*-values come from. When students have a notion of distribution that is not based on visual aspects (e.g., center, spread, skew), the students can, with support, come to think about distributions as entities that have characteristics that we can measure. Thus, students have the opportunity to engage in the act of quantification (Thompson, 2011) of these characteristics. Often, *p*-values are introduced to students as “the probability that we get the value we got or one more extreme” and *p*-values are “areas under the curve”. This approach makes *p*-values something the student finds from a graph and hinges on a quirk of the Cartesian coordinate system rather than allowing the student to make the connection between a *p*-value and the output of the cumulative density function.

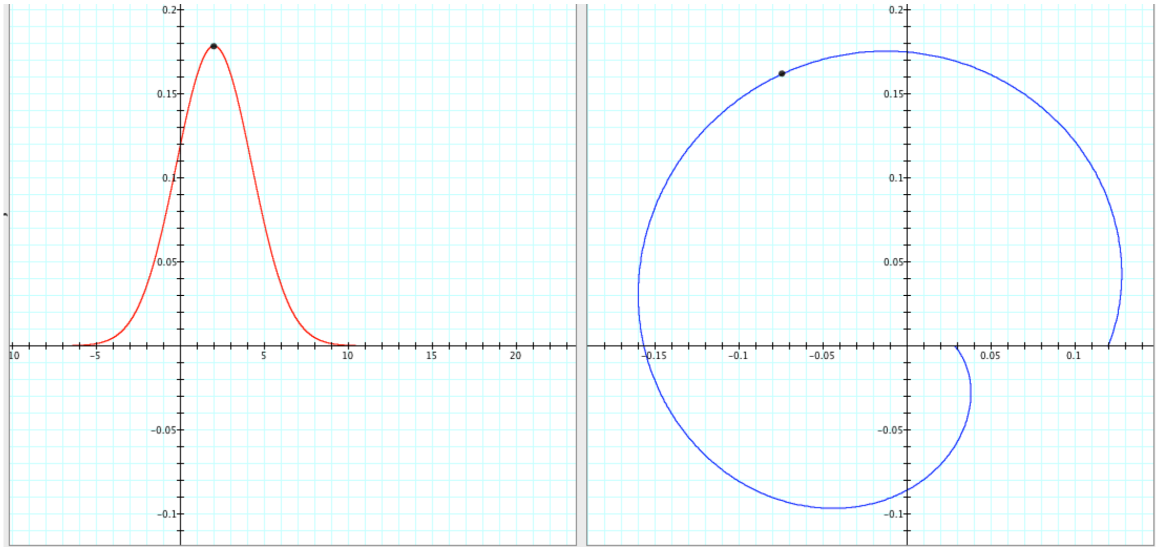


Figure 15. Shortcoming of *p*-value as "area under the curve".

Suppose that we have a stochastic variable that is cyclical (such entities do exist; see Dwass, 1962). If students operate using standard shapes and approaches, the student will attempt to find area under the curve in the graph (of the PDF) to the left. However, if the student embraces the cyclical nature of the random variable, she will produce the graph (of the PDF) on the right. The notion that a *p*-value as the "area under the curve" breaks down. (The black dot identifies the point on the PDF that corresponds with the mode of the distribution; both graphs are of the same PDF.) Providing students with the opportunity to connect *p*-values with outputs of cumulative density functions (a small extension from my proposed target meaning for distribution) with a potentially stronger basis for them to build powerful ways of thinking about hypothesis testing (parametric and non-parameter) that underpins all of inferential statistics.

Chapter 5: Methodology

This dissertation study seeks to address the following research question: "What meanings for stochastic process do students develop during an instructional sequence based upon a hypothetical learning progression for conceiving of distribution as describing the complete behavior of a stochastic process?" Looking at students' meanings for related fundamental concepts including randomness, random variable, and probability supports this central research question. The purpose of this research is to develop a working theory for students' understanding of one of the core concepts of statistics. This purpose is in line with the goal of design experiments (P. Cobb, Confrey, Lehrer, & Schauble, 2003). Cobb, Jackson, & Dunlap (2016) identify five feature of design studies: 1) a focus on generating theories for how students' learn, 2) an interventionist nature, 3) a theoretical and pragmatic orientation, 4) iterative design, and 5) generalizability.

For this study, I propose to recruit students who completed an introductory statistics course for the life sciences during the Fall 2016/Spring2017 semesters. I chose this course for several reasons. First, most students taking introductory statistics courses are non-math/stat majors, so I wanted to work with a non-majors population. Second, I've taught this course several times, each time refining the activities that underpin this study. This allows me to draw on the iterative design feature of design studies and now more closely examine what understanding students have from the sequence of activities that I used in Fall 2016. Several activities have roots in the Fall 2016 course and have been refined during the Spring 2017 semester in another statistics course intended not only for mathematics/statistics majors but also for non-mathematics graduate students needing an introductory course. Third, this course allows me the opportunity to draw potential students from several sections that did not focus on trying to support students in constructing the target meaning of distribution and stochastic process. This will allow me the opportunity to see what meanings these students have at the end the course and what struggles they might have in developing the target meaning when working through the sequence of activities. As Tufte (2006) notes comparisons are a fundamental act in statistical reasoning. By comparing students who went through early versions of the activities in the class to students who did not, I'll be able to get a better idea of the various meanings that students might have for the concept of distribution. For students willing to participate, there are two distinct phases; a clinical interview and the instructional sequence spanning three activities.

Phase One

In the first phase, students will participate in clinical interviews (Goldin, 2000). These interviews serve as a way to characterize the meanings that each student currently has for randomness, random/stochastic variable, random/stochastic process, probability, and distribution. Examples of these questions appear in the Appendix A. These many of these questions come from the existing research, allowing me to place students amongst those described in the related literature. This affords me a strong theoretical and pragmatic link between my research and previous research as well as laying the foundations of generalizing my research. The clinical interviews will enable me to not only characterize individual students but to create matched groups of students for the second phase.

Phase Two-Light Switch Activity

The second phase of the study is the primary intervention phase. As Cobb et al. (2003) note design research can be done at several levels from one-on-one settings to entire classrooms. After the clinical interviews, I'll explore what meanings students develop and the difficulties have while working through an instructional sequence. In particular, I will use a one-on-one teaching experiment (Steffe & Thompson, 2000) to build models for how the students are thinking. Students will work through an instructional sequence (described below) based upon the conceptual analysis from Chapter 4. I will form hypothesis as to each student's meanings to build testable models for how the students are thinking. These hypotheses will allow me to revise any construct maps, the learning activities, and the learning progression. At the onset, I'll use the following sequence of tasks to support students in developing a meaning for distribution as the accumulation of stochastic process outcomes with respect to the value of the stochastic variable. These tasks are the results of several design/testing/revision phases that began with a course project[6] and continued through four semesters of teaching. The students who worked through were predominately majoring in the life sciences and were taking an introductory course in statistics.

During this phase, students within each matched group will be assigned to one of the two versions of the Light Switch Activity sequence. These two versions will enable me to look for affordances and hindrances of the instructional sequence on student learning. The questions/tasks posed to the student will be the same in both versions. The difference will be the sequence in which students work. In version A, students will see

[6] My thanks to Gabriel Tarr for his assistance during the initial development.

all four rooms to the activity and move from deterministic switches with increasing variation to a fully stochastic switch. In version B, students will see one room at a time and start with the fully stochastic switch. Should a student in version B struggle, they will be shown other rooms in the reverse order of version A. A description of version A follows.

Light Switch Activity-Version A. The first activity focuses on helping the student make a distinction between a deterministic process and a stochastic process. Figure 16 shows the user interface for the Light Switch applet[7] that serves an object for discussing processes. The interface consists of a screen with two grey lines sectioning out four "rooms" as well as five buttons. The "Setup/Reset" button is for the researcher to activate the app. The student will be asked to work with the other four buttons, all labeled as "Switch". These four buttons are what the student will interact with in the applet as he/she works through the activity.

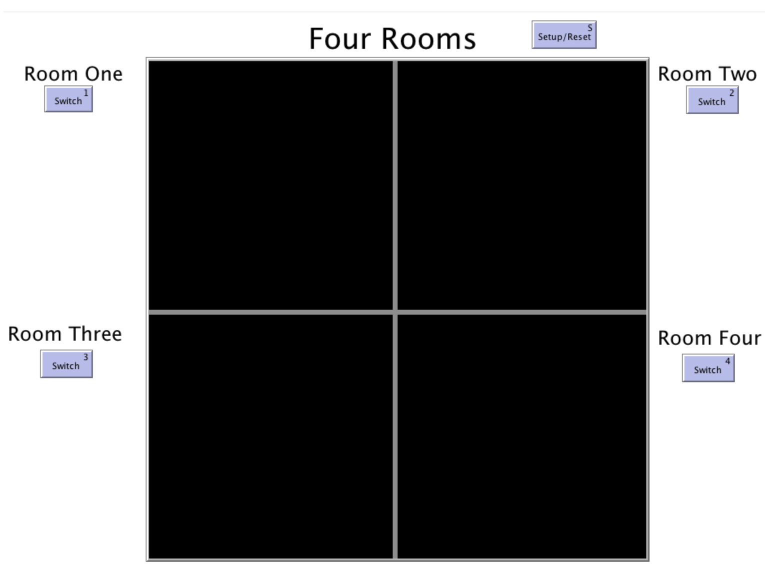


Figure 16. Light Switch applet interface.

---

[7] A web version is available: https://mathpost.asu.edu/~neil/dissertation/FourRooms.html.

At the start of the activity, students are asked what is the first thing they do when entering a dark room. This question primes them exploring the behavior of light switches. Room One's button/switch is a deterministic process with two levels; the room turns white when the switch is "on" and black when the switch is "off". (All rooms start with the light out.) Room Two's button/switch is also deterministic but acts like a dimmer switch; the first five clicks move through lightening shades of grey until the room is fully lit (white) on the fifth click. Starting with the sixth click, the room darkens until the reaching full black on the tenth click. The grey colors used in turning on the light and turning off the light are the same. The intent behind Room Two is to start introducing more variation in the results of the deterministic process and see what students make of this variation.

Room Three plays off of the increased variation in Room Two but in a different way. While the switch here is also deterministic, the light in the room cycles through the colors red, orange, yellow, green, blue, indigo, purple, magenta, pink, and white with black ("off") occurring between each color. There is more variation to this process than in the previous two not only in number of colors but also in the inclusion of non-grey scale colors. Room Three's switch follows a fixed pattern to the colors (they occur in the order I listed), making the process deterministic.

This brings the student to Room Four. The switch in this room is a stochastic process where the color that appears starting with the first press of the button follows a heavily modified continuous uniform distribution. Unlike Room Three where there is a defined "off" between each color, the light in Room Four does not necessarily turn off with every even numbered click; rather the room's light will only go out if one of the color codes for black return. Making use of uniform distributions is a common practice in research literature for multiple reasons. From a technical standpoint, stochastic processes with uniform distributions are the easiest to program. From a pragmatic side, students have experiences with uniform distributions thus they have a better chance developing an understanding of the process that he/she can explicate. However, I must stress my usage of the phrase "heavily modified". The colors that appear do not actually follow a continuous uniform distribution. Rather, the applet will generate a number from a continuous uniform distribution on the interval [0, 140). The applet then rounds this value to one decimal place and then looks up the color associated with the rounded value. The color palette within the applet system maps 14 unique values to the color black as well as 14 other unique values to the color white; there are an additional 28 unique color codes that map to near-black (as well as near white) that are visual indistinguishable from black or white. Taken together, the rounding and the color palette ensure that stochastic process between Room Four's button/switch truly stochastic with limited knowledge of the actual distribution. Figure 17 shows examples of the lit rooms in the applet.

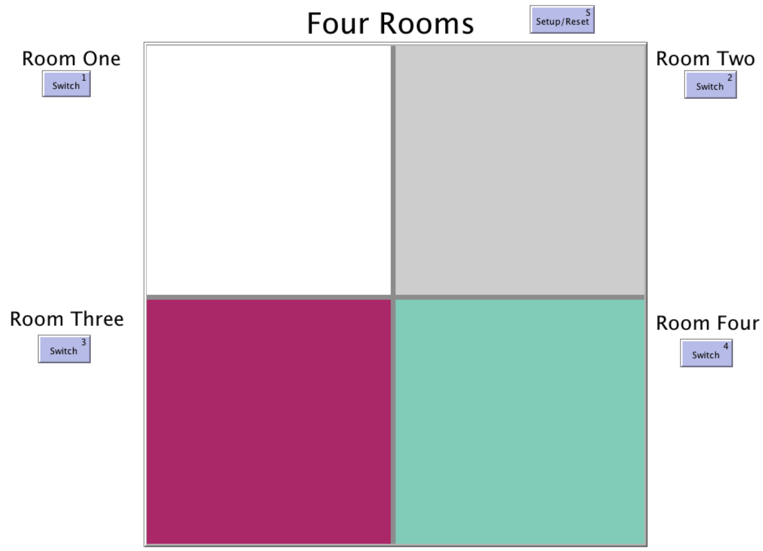


Figure 17. Light Switch applet states for Version A.

Light Switch Activity-Version B. As previously mentioned, Version B of the Light Switch Activity presents one room at a time (see Figure 18). Students start with a room whose switch is identical to that in Room Four of Version A. When a student has problems answering questions about the current room, I will close out the room and open up a new room in a new window of the program. This new room will be equivalent to the Room Three in Version A. Each time the student runs into problems, I will back up another room. If a student resolves his/her difficulties, we'll return to the previous room and see if the student can now answer the question(s) he/she originally struggled with.

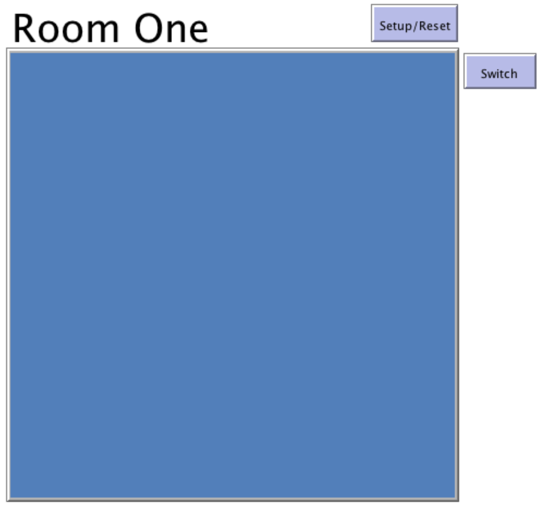


Figure 18. Light Switch applet, Version B

Questions for the Light Switch Applets. Both versions of the Light Switch activity will involve the following questions:

- Press the button/switch for the room. What happened?
- Press the button/switch for the room again. What happened?
- What do you think will happen when you press the button/switch again? Twice? How certain are you? What makes you certain?
- What is happening when you press the button/switch?
- [From the current state of the room, I press the button to change the state] Do you think that it is possible to return the room to state it was just in? What would a person need to do? Would that method always work?
- Suppose that current switch broke, and you needed to make a new one. Describe the rules that you make for the new switch so that room continues to function.

- Are there any similarities between the rooms? Any differences? (Only in Version B if the students see more than one room)

These questions serve several goals. The first questions help the student to familiarize themselves with the room's switch. The middle questions seek to transition the student from talking about the outcome of the process to the underlying process, culminating in the student needing to generate his/her version of the process. The last question is meant to get the student to think about all of the rooms build comparisons that he/she can use to decide whether new processes are deterministic or stochastic (or whatever labels he/she generates).

Phase Two-More Processes Activity

The second activity of the second phase is the More Processes activity. Regardless of which version of the Light Switch activity the student saw, all of the students will see the same things during this stage. The intent in this phase is to see what the students carry with them from the previous stage. The More Processes activity presents the student with five new processes for the student to reason about; see Figure 19. Unlike the processes involved in the Light Switch activity, the student does not necessarily have an interface with which to carry out these processes. The student must now imagine carrying out the processes. The sole exception is the Neil's shoes; this process can be carried out by the student but only in limited form. The processes that appear in this activity reflect a mix of deterministic and stochastic processes as well as process that

Suppose that 20g of baking soda ($C_2H_4O_2$) are combined with 30.2g of vinegar ($NaHCO_2$) at room temperature (70°F). [Creates 4g of water (H_2O), 9.77g of carbon dioxide (CO_2) and 36.43g of sodium acetate ($NaC_2H_3O_2$)]

Suppose that everyone in a class writes his/her name on separate, identical Ping-Pong balls and then places his/her ball into a bag. Once everyone has, the teacher shakes the bag and reaches into the bag and removes one ball to see what student wins $10.

Suppose that we place $1000 into a savings account with Barclays at an APY of 1.00%. We do not deposit any more money, nor do we withdraw any. We check the account balance once a year, every year.

Suppose that we stand at the intersection of Rural Road and Apache Boulevard from 2:00pm to 4:00pm every Monday and record the total number of red vehicles that go through the intersection.

You look at Neil's shoes and record the most prominent color.

Figure 19. The new processes in the More Processes activity

could go either way (Neil's shoes). For the stochastic processes, there is a first-order processes (lottery) and a second-order/time series process (counting red cars). While the lottery is a familiar process for most students, this stochastic process is similar to that in the Light Switch activity. The counting of red cars is a type of process that the student may not have much familiarity with. The Neil's shoes process sits in an interesting position. As posed to the students, there are sufficient details missing that students could view the process deterministically or stochastically. Given the student's initial classification, I intend to probe the student to see the image of the process he/she has constructed for this process. However, at a certain point, I'll ask to student how he/she might re-conceive this process so that it would fit the other category. This line of questioning should help me to further characterize what the student sees as the distinctions between deterministic and stochastic processes. Once the student is comfortable with the new processes, I will ask him/her to group these processes into as many categories as he/she feels is necessary, probing what each category represents and why each process is in that category. I will then ask the student where he/she will add in the room(s) from the Light Switch activity to his/her categories.

If a student struggles to form categories during this activity, I will enact a second intervention. This second intervention stems from an activity I conducted in the Spring 2017 iteration after the students reached an impasse in which of their two categories of processes the counting of red cars belonged to. The nature of their comments revealed that the arguments revolved around their images of outcomes, repetition, and the rules/details of the processes. I'll draw the student's attention to these features with discussion of what these features consist of and how each feature varies. Figure 20 shows an organizational table that can be used with the students during this intervention.

Process	Fixed or Unfixed Outcome?	Type of Repetition: Reproducible or Replicable	Clear or Fuzzy Rule?	Category
Room One				
Room Two				
Room Three				
Room Four				
Baking Soda + Vinegar				
Labeling Ping Pong Balls				
Savings Account				
Counting Red Vehicles				
Looking at Neil's Shoes				

Figure 20. Second intervention organizational table for students

Phase Two-Sequences Activity

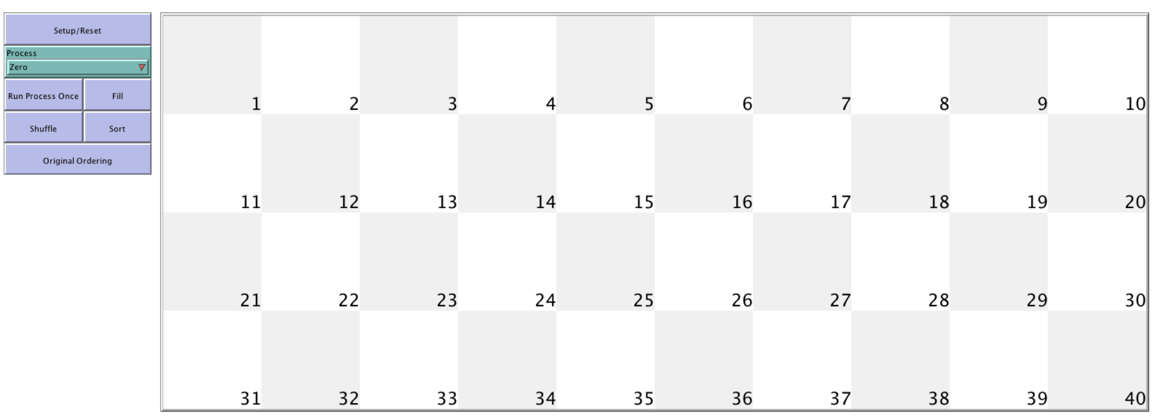


Figure 21. User interface for the Sequence applet; showing Process One's full sequence.

The final activity in Phase Two is the Sequences Activity[8]. The activity still asks the student to classify a given process as either deterministic or stochastic. However, unlike the previous activities, the processes are hidden from the student and he/she only has access to the results of carrying out each process a total of 40 times. This activity simulates the standard practice of using a sequence of outcomes to classify the generating process as being deterministic or stochastic. This activity's constraints help to reveal the features of a student's meaning for stochastic process that he/she places a premium on. Figure 21 shows the Sequence applet interface used for this activity. In total, I've programed fourteen sequences into the applet. The first sequence ("Process Zero") is a simple deterministic process of adding one so that the student can get used to the interface. There are eight fully deterministic processes, four fully stochastic processes, and two processes that are open to interpretation (Processes Two and Seven) as listed in Table 8.

Process Two is such that a student could view the sequence as coming from either a deterministic or a stochastic process, but the two images are distinct. For example, if the student imagines that Process Two is a listing of nucleotides of a single person's DNA, then the sequence is deterministic. However, if the student instead imagines that the sequence is the listing of the first nucleotide in separate individuals' DNA, then the process is stochastic. Process Seven has deterministic and stochastic elements. Using von Mises's Principle of the Impossibility of a Gambling System would make this process non-stochastic. However, how a student thinks about such a process is unclear.

---

[8] A web version is available: https://mathpost.asu.edu/~neil/dissertation/Sequences.html.

Process Thirteen is a stochastic process that is a stationary time series. This provides a similar process to the counting of red vehicles from the More Processes activity.

Table 8. The Fourteen Processes in the Sequence Applet

Process	**Rule, (let *n* represent trial number)**
Zero	n
One	Each integer listed in increasing order starting with 1 but each prime integer is repeated the same number of times as that prime's value.
Two	One of "A", "G", "C", or "T" using $\mathcal{DU}(4)$
Three	$2n + 1$
Four	$n^2/100$
Five	A value from $\mathcal{N}(0,1)$, rounded to two decimals
Six	Return a value from $\mathcal{DU}(6)$
Seven	2 if $n \equiv 0 \bmod 3$ A value from $\mathcal{DU}(5)$ on the set {1, 3, 4, 5, 6} otherwise
Eight	$n - \sin(45)$, rounded to two decimals
Nine	-6
Ten	Fibonacci's Sequence, rearranged in blocks of four
Eleven	A value from $\mathcal{E}xp(2)$, rounded to two decimals
Twelve	$-1/n$, rounded to five decimals
Thirteen	$X(n) = X(n-1) + err(n)$, where $err(n) \sim \mathcal{N}(0,25)$

As the student works through each sequence, I will ask him/her for what she thinks is going on and to develop a description of the process similar to what he/she did with the light switches and saw with the More Processes activity. I will then ask him/her to classify that process into one of his/her existing categories. This activity will help me further characterize each student's meaning for stochastic processes.

This sequence of activities and interventions should allow me to describe the meanings that students have and develop for stochastic processes. The stochastic processes embedded in the activities cover not just familiar ones (e.g., uniform processes) but also non-uniform first-order stochastic and second-order stochastic processes. This variety helps to ensure that the meanings that I characterize aren't necessarily biased by my sole usage of stochastic processes of a single type. If students routinely fall back on heuristics such as equiprobability, then this provides useful information regarding elements of their meanings.

Chapter 6: Students' Meanings for Randomness

To understand stochastic processes is a necessary condition for understanding and using any type of statistical distribution. In thinking of distribution as the accumulation of outcomes from a stochastic process, an individual must imagine a process that generates these outcomes. However, the way that the individual conceives of the process must entail her imagining that the process has unpredictability in short-run, predictable in the long-run, and having few sources of bias. In other words, she must view the process as being random; that is, her imagined process must by a stochastic (or random) process. In trying to look at how an individual might think about stochastic processes, looking at her meanings for "randomness" become vital. In this chapter, I will describe the main questions of the clinical interviews intended to get at students' meanings for randomness, describe how three students appear to think about randomness, and discuss the implications of these results.

Clinical Interview Questions on Randomness

As part of the clinical interview, students were asked to respond to two primary questions dealing with randomness. The first question (Figure 22) asked students to select which of eight situations/statements matched his/her meaning for randomness. The second question asked students how they would explain the idea of randomness.

The first question's situations/statements correspond to different levels of the construct map for randomness (see Table 5); situation G corresponds to the Sequence Attribute (the highest) level, with situation F corresponding to the second highest level in the map (Sequence Complexity). Situation A links to the Left-field level while situation B to the Unknown/Unpredictable level. Situation C can go to either the Left-field or

Unknown levels, depending on how the student interprets the situation. Situations D and E link to the lowest levels of Ordained and Chaos, respectively. These situations provide opportunities to explore how students are conceiving the situation and their images for randomness. Later in the interview, students were asked how they would explain the concept of randomness.

Select all of the following situations that you believe match your meaning for "random".

A) Tom and Harry are in the break room discussing what they thought about Star Wars: The Force Awakens. While describing what he liked about the movie, Tom said "Oh, did you know that Linda (a co-worker) is Lutheran?" Harry replied, "That's random." [Left-field]
B) You're at home, someone knocks on your door and you don't know who it is. [Unknown/Unpredictable]
C) You and your two closest friends are trying to resolve who gets to choose what movie to see. One friend doesn't care but the other one and you both want to go see different movies. The neutral friend picks a number at random and the closest of the other two friends wins. [Left-field, Unknown/Unpredictable]
D) Nothing is ever random; there is always a reason that things occur. [Ordained]
E) Everything is random. [Chaos]
F) A sequence is random when you can't find a pattern to it; like the number pi. [Sequence Complexity]
G) A sequence is random when you can't find a pattern, but you can use it predict something in the long-run. [Sequence Attribute]
H) None of these match my meaning for "random".

Figure 22. First clinical interview question dealing with randomness

Descriptions of the Three Students

For this chapter, I'll discuss three students, Bonnie, Colin, and Danielle; all three students took an introductory statistics course geared towards students majoring in biological sciences in the Fall 2016 semester (one year prior to interviews). Bonnie switched her major to elementary education while Colin focused on genetics, cell and developmental biology, and Danielle majored in conservation biology and ecology. In addition to the introductory statistics course, Bonnie had taken pre-calculus in high school, a college mathematics course in Spring 2017, and was currently enrolled in a mathematics for elementary teachers course. Danielle had taken intermediate and college algebra as well as an online precalculus course in addition to her introductory statistics class. Colin differed from Bonnie and Danielle in his mathematics course work. Colin had completed a full sequence of calculus (three courses) as well as a differential equations course; he took his introductory statistics course while enrolled in these courses. During the Fall 2017 semester (when the interview took place), Colin was enrolled in an introductory proof course as well as a second introductory statistics course designed for mathematics and computer science majors. He cited his enjoyment of his first introductory statistics course as a reason for considering adding a Statistics minor.

Bonnie and Danielle also differed from Colin in another way; both of them took the introductory course from instructors who followed a procedural and traditional approach to introductory statistics. Colin, on the other hand, had a reformed approach from this author that focused on helping students build productive meanings for statistical concepts. While Colin was exposed to pre-cursors of the activities in this study, the activities used here are far removed from what he saw then. However, Colin's second introductory statistics class did mark a return to procedural statistics. Bonnie received a B+, Colin an A, and Danielle a B in their first introductory statistics courses.

Bonnie's Meaning for Randomness

From the start of the interview questions dealing with randomness, Bonnie brought up a connection to one thing having an impact (or lack thereof) on the outcome/result of some event. This is evidenced by Bonnie's description of what first came to her mind after reading the first question dealing with randomness as evidenced in line 1 of Excerpt 1:

Excerpt 1. Bonnie's response to "random" situations

1	Bonnie	I think that if something is random, it doesn't affect the outcome.
2	[edited]	[~1.5 minutes later]
3	Bonnie	So, when I first read these I started recalling "randomness" from Statistics and I know that something's random, I don't know if I have it switched around but, if it is or isn't, it does or does not affect the outcome, I forgot which one or which way, but...actually, any of this shouldn't affect any outcome unless, hmmm...okay, I think it is not random because it, by saying that oh, this person is Lutheran and it could affect of, it could affect someone's opinion about them. I'm not sure (laughs)
4	Neil	And why did you pick B (Unknown/Unpredictable) as being something that matched your meaning for randomness?
5	Bonnie	Umm,...I picked B (Unknown/Unpredictable) as random because unlike A (Left-field) you're not in a situa-, you're not in a conversation talking about a specific item, it is just something that happened with no pattern, no sequence, or anything
6	Neil	So the idea is this sort of an out-of-the-blue [Bonnie: Yeah] type things. Umm, you talked, you talked a little a bit about random as either effecting an outcome or not effecting an outcome [Bonnie: mmm-hmm (agreement)], would that apply here in B (Unknown/Unpredictable)?
7	Bonnie	Yes, it actually would. Depending on who the person is.
8	Neil	The person...?
9	Bonnie	That's knocking on the door.
10	Neil	But we don't know who [Bonnie: Yeah] Does that make you question yourself on whether or not you think that B matches your sense of randomness?
11	Bonnie	Yeah it does (laughs) Umm, ... so, ... I don't know if it is random any more since it will affect the outcome cause you don't know who it is, it might be someone that's delivery mail, it could be someone that's trying to kill you, I don't know, (laughs) but I think it could affect...the person who is sitting at home...just a random person knocking on the door...(softly) I think...(laughs)

Bonnie's initial recollection of randomness as being tied to affecting the outcome could be potentially grounded in the traditional presentation of using randomization to minimize sources of bias and other confounding factors on the results of a study (line 1).

The traditional approach of presenting randomization as a way to minimize sources of bias provides a route for students to connect to the idea of the independence of events. However, Bonnie is unsure of whether randomness indicates that there should or should not be an effect on the outcome (lines 3, 7, and 11). I asked her to clarify which she thought in line 6; Bonnie started shifting away from being uncertain towards that randomness should impact the situation. This is further evidenced in Excerpt 2:

Excerpt 2. Bonnie solidifies her meaning for "randomness"

1	Neil	You seem a little hesitant now [Bonnie: Umm-hmm (agreement)] So you, are you hesitant about your meaning for randomness?
2	Bonnie	Yes (laughs)
3	Neil	Umm, cause we can keep going down the list with this same idea, umm, cause...Have you come to a decision: randomness should not affect outcome or randomness should affect outcome?
4	Bonnie	Umm, let me think for just second, so I'm thinking that if something's random so if I just, as an example, I'm randomly picking a nail polish color will that, that could affect the outcome, or someone randomly picks a name from a hat that could affect an outcome of, I don't know, if you are presenting so I think that randomness does affect almost everything.
5	Neil	So, how do you feel about C [left-field] now? You initially said was not random.
6	Bonnie	Umm,...,so I think it is random now. Because whichever number you choose the person who is closest to that, they're going to have a completely different outcome if they choose a scary movie compared to the other friend who wanted to choose a funny movie [Neil: Okay] So then two different outcomes.

Bonnie's second example (drawing names) is a common example used in introductory statistics classes. In line 4 of Excerpt 2, Bonnie once again mentions how randomness "affects almost everything". The notion of randomness impacting or affecting something is a reoccurring theme for Bonnie. The following excerpt (Excerpt 3) entails how Bonnie would explain the idea of randomness:

Excerpt 3. Bonnie’s response to explaining “randomness”

1	Neil	How would you, so earlier I had you sort of look this list and pick the ones that you thought matched your meaning for randomness [Bonnie: umm-hmm (agreement)]. How would you explain the idea of randomness?
2	Bonnie	Okay. So, for randomness...(long pause) I want to go back to the term "everything is random" but in reality, it's not I want to say but...for example, it wasn't just for me to be in this research study wasn't random, you had to look through if I was in a specific class or not and maybe I just happened to respond to you compared to other people or students that took statistics, so random by that. Ummm, umm, so...random is, I...I believe it would be the outcome, err, how something would affect a long-term run almost, of a situation.
3	Neil	So what do you mean by a "long-term run of a situation"?
4	Bonnie	Umm, just how randomness can affect on a situation. Going back to the example of the movies where by picking a random number, whoever is closest to it, well that can affect the long-run if they want to see a scary or a funny movie. One person could be sitting there being all sad and upset the whole time while the other person enjoying the movie...So, (laughs) this is hard...
5	Neil	So what are you thinking about right now?
6	Bonnie	I'm thinking about...not the actual definition of randomness, but how it affects everything in the situation
7	Neil	How randomness or the definition?
8	Bonnie	How randomness can affect a lot of things in a situation. But I'm trying to think of a definition but...(laughs)

Bonnie mentions randomness affecting things in lines 2, 4, and 6. For her, what is being impacted lies beyond the immediate result of the process (e.g. the result of the coin flip). When she mentions that “situation” and “long-run”, Bonnie describes how she imagines the described movie situation playing out after a choice is made. The friends go to a movie and the friend who lost being sad and upset (Line 4). Bonnie’s introduction of the emotional states of the individuals in the movie situation (situation C) helps her see how individuals are affected by things that are random. The “randomness” of the situation is impacting the people in the situation and that is what Bonnie appears to focus on rather than whether or not whoever was closest (the actual result of the process) was impacted.

While Bonnie sees the potential for various outcomes in situations A-C, she is does settle on the notion that randomness should impact the situation via the different outcome that occurred. I then asked her about applying her meaning to situations D (nothing is random), E (everything is random) along with situation F (sequence with no pattern, like π).

Excerpt 4. Bonnie’s view of nothing/everything being random and sequences

1	Bonnie	[after reading statement G-sequence attribute] Umm…I don’t think that G is random. If you’re going to predict something that will affect another item in the long run, then it shouldn’t be random.
2	[edited]	[~4.5 minutes later; see Excerpt 1 and Excerpt 2]
3	Neil	So how about D?
4	Bonnie	...So, I...I think that...there is always a reason, right (sigh), okay, I'm confusing it with E right now.
5	Neil	So, if ignore everything else, just look at statements D [ordained] and E [chaos] [Bonnie: mmm-hmm (agreement)], what's the relationship between the two of them?
6	Bonnie	Umm, well, if you're saying that nothing is every random and that there's always a reason that things occur for D and E is saying that every is random and from my understanding, whatever you, if something is random there's still going to be an effect or reasoning [Neil: umm-hmm] behind that, so I that the reasoning will come after the randomness and that, for why it occurs to begin with
7	Neil	Okay
8	Bonnie	Umm,...,(laughs)
9	Neil	Do you still disagree with both statements?
10	Bonnie	Umm...I...agree with E...,[long pause]..., and then for D...I...disagree with D, but I agree with E
11	Neil	Okay. And then you said that F was also a match for your meaning for randomness, is that still true?
12	Bonnie	I believe that it is.

Bonnie initially struggled with what situation D and E were asking (lines 4 and 6 of Excerpt 4). Given that Bonnie seems to think about randomness as how the result of some action will impact the situation (e.g., friends' emotional states, color of nails, etc.), her agreement with the statements of "Everything is random" fits within that model.

While Bonnie appears to view randomness as whether something impacts a situation, she separates out the reasoning behind that impact into pre-event knowing and post-event knowing. I hypothesize that Bonnie's shift from her initial view of randomness as not impacting outcomes was shifted by her trying to reconcile her decisions about situations A-C and F as all matching her image of randomness. We can glean insight into Bonnie's response to statement G (line 1 of Excerpt 4) by looking at line 6. For Bonnie, knowing the reason behind the event before the event occurs is grounds for the lack of randomness. Bonnie might view a prediction as being predicated on know why something occurs and therefore affecting the outcome in a different way than her imagery about randomness affecting the situation. I believe that Bonnie was struggling to separate two distinct aspects of what the process was affecting. She thought of an immediate result of the described process (of which she could generate several examples) and how that result would play out in the given situation with the imaginary individuals. This is to say, she imagined sample spaces for the processes and emotional reactions to each member of those space. This can be found when Bonnie brought up the idea of randomness during a different interview question about picking a machine to model the flipping of a fair coin (Excerpt 5). In line 5, Bonnie indicates that the effect of the randomness is what you do to record the result of a single trial.

Excerpt 5. Bonnie's brings up randomness when discussing coin flipping machines

1	Bonnie	When you do toss a coin [Neil: umm-hmm],..., I'd believe that Machine 3 is correct, cause it gives you all of the possible outcomes.

2 Neil So is that the only machine that is correct?
3 Bonnie Umm,...,umm,...,I mean I guess that they could all be correct. cause it just happens by randomness I guess...
4 Neil So what do you mean by that statement?
5 Bonnie Umm, if you're flipping a coin, it's random but the outcome will still affect...what you...put down in the sequence, hmm...I still, uh, I think that they are all correct but I think that Machine 3 is the best model to have.
6 Neil Okay...So if you had to pick any one to flip coins for you, you'd pick Machine 3?
7 Bonnie (softly) I think so.

Bonnie agrees that unplanned, her “cluster of craziness”, would be a synonym for randomness to close out Excerpt 6 (lines 2-4).

Excerpt 6. Bonnie’s synonyms for randomness

1 Neil So, with the last question, you gave synonym [Bonnie: umm-hmm (agreement)] for "chance", what might be synonyms that you would use with randomness?
2 Bonnie Ummm, (long pause), I...I think of antonyms which would be [Neil: okay] "chosen" or...hmmm,..."recommended", a recommendation,...or "specific". [Neil: Okay] So I really hold these words are pertaining to something specific you're looking for rather than randomness is...when I think of randomness, I just think of a cluster of craziness and you'll never know what you pick out...(long pause)
3 Neil Would you say that the word "unplanned" [Bonnie: yeah] would work as a synonym?
4 Bonnie Yes, I would say that or so, somewhat spontaneous...
5 Neil Okay...Any other thoughts with randomness?
6 Bonnie Umm, (shakes head and laughs) [Neil: Okay]

As a whole, Bonnie’s conveyed meaning for randomness deals with whether the result of a single trial will impact the larger situation. While she initially invoked the idea that randomness should not affect the outcome, Bonnie’s focus shifted from the outcome (the result of a trial) to the situation she imagined as continuing forward. For her, randomness should impact what would then happen in the hypothetical situations. The possibility of branching points along with not knowing a reason pre-event appear to be hallmarks for Bonnie to declare whether something is random or not. Her conveyed meaning becomes a driving force for her in a later question about her meaning for a random variable and her giving an example of one (Excerpt 7). Her example (line 4) appears to stem from her conveyed meaning of many possibilities that impact the larger situation.

Excerpt 7. Bonnie's example of a random variable

1	Neil	Could you give me an example of a random variable?
2	Bonnie	Hmmm, (long pause), I'm trying to think right now and I can't
3	Neil	It's okay; remember there are no wrong answers
4	Bonnie	Yeah. So earlier I defined random as something that most of the time will always have an effect on the outcome...so if you picked out of a hat and...you picked either one, two, three, or four, or five and whatever number you got was the order you went to present. Well, if I got a three that's a random variable but its gonna have an effect on how I'm gonna, how nervous I'll be to present or how calm you'll be to see [Neil: Okay] So, yeah, I think that's my answer.

Colin's Meanings for Randomness

Early on in talking about the concept of randomness, Colin brought up a distinction between what he saw as two separate meanings for "random"; one that is colloquial and one that is technical (Excerpt 8, lines 2 and 3).

Excerpt 8. Colin's initial response to "random" situations

1	Neil	Here's a whole list of situations and I'd for you to read them each aloud, in turn, and share me, with me, your thoughts about each one, and whether or not you believe that situation demonstrates your meaning for randomness.
2	Colin	Okay, So, Tom and Harry are in the break room discussing what they thought about "Spiderman: Homecoming". While describing...while describing what he, uhh, liked about the movie, Tom said "Oh, did you know about, did you know that, uh, Linda, a co-worker, is Lutheran?" Harry replied, "That's random".
3	Colin	Umm, so, because of a prev- previous Statistics classes, I don't think "random" as a, uhh, I don't think of "random" in the way of

		"oh that doesn't follow the continued line of thought", I think of random as a, more kind of mathematical side where you cannot find a pattern or you cannot, ahh, predict what in a sequence what the next value will be. That's what I consider as random. So, in this case, umm, I would say, that that doesn't fit my, uhh, definition of something being random.
4	Colin	Now I do know, before I took Statistics classes I would use "random" in kind of the context they did there where it's like "oh, that doesn't follow the line of thought; that doesn't make sense" kind of thing, so you would say that's random, but anymore because of Statistics class-my Statistics classes, uhh, I don't think in, I don't connect random with that process anymore.

Colin's non-mathematical meaning is consistent with what I called "Left-field" randomness (see Table 5). His mathematical meaning initially appears to be consistent with my "Unknown/Unpredictable" category however, he quickly adds on the stipulation that there must be a sequence of occasions rather than a single instance in Excerpt 9.

Excerpt 9. Colin's mathematical meaning for "randomness"

1	Colin	So in B [unknown/unpredictable], You're at home, someone knocks on your door and you don't know who it is.
2	Colin	Umm...the following that are random...umm. Yes, in the way that you can't predict who is gonna, who the next, I guess, person in that series is going to be. But the thing is no previous series, so umm, I don't know if you could establish something as being random and that, using a kind of a mathematical approach. Umm...
3	Neil	So it sounds like B is sort of "iffy".
4	Colin	Kind of "iffy" but I'm still leaning on the "no" side because you're only looking at one, uhhmm, one instance there and you can't really say that something is random with an observance of one. [Neil: Okay] Ummm…
5	Colin	So, C [left-field], you and your two closest friends are trying to resolve who gets to choose what movie to see. One friend doesn't care but the other picks, ahh, one doesn't care but the other one...ahh,
6	Neil	The other one and you…
7	Colin	so, yeah, oh! because it is two friends, kay yeah. One friend doesn't care but the other one and you both want to see different movies. The neutral friend picks a number at random and the closest person wins. Okay, so that, that one is a little closer to what I would define as "random" than the other two examples, ahh-umm,

because the two people choosing numbers, they don't know what number that has been, what number is next in the sequence and their, ah-umm...it's closer, but the only thing is probably still, like, if they say a number between 1 and 10, you are still limited to ten possible outcomes there. Ahh-umm, so...(to self) would that be random? I would say that's...iffy but border side "yes". Uhh, that would be a more, more considered, what I would consider random.

In line 4 of Excerpt 9, Colin dismisses the event of not knowing who is knocking on the door as being random as there is only a single instance. When he moves into the three friends situation, he conveys that he's thinking of a sequence of numbers. One possibility that could explain why Colin envisioned a sequence of numbers but not a sequence of people: in his experiences, he might have worked with a random number generator and/or a table of random numbers. Additionally, he might have drawn upon his own experiences in such a number-guessing game. In both of these cases, Colin's imagery allowed him to more readily envision the described process repeating. For the person knocking at the door, Colin's imagery is grounded in a non-repeating event. We can see how his image of a sequence interacts with his meaning for randomness by looking at how he explains whether nothing or everything is random in Excerpt 10:

Excerpt 10. Colin explains whether nothing or everything is random

1	Colin	D [Ordained]. Nothing is ever random; there is always a reason that things occur. Philosophical there. Ummm, well, I mean, there are random things like for example, the motion of like atoms can be considered to be highly random, but they are influenced by some outside force. So, is that considered random? Umm, I guess for me it just kind of looks on how, how deep would you be looking into it. Umm, kind of like if you were to, I guess on a higher, like on a higher level, it would, yeah, they would look random, but if you then started looking at like the interactions between the atoms or whatever, you would start saying well maybe they aren't, they not so random.
2	Colin	So I guess I would compare that to if you were looking at a sequence and once you see a large enough sequence, you have a

large enough, uh, sequence or array of data, then you can start pulling out patterns if they are and then you can say "oh that's not random" but if you still can't see any pattern then you can say that it's not random [misspeak?]. Unless, of course, if you keep zooming out, or you keep drawing a larger and larger, uhh, set of data, at that frame of reference, and maybe see a pattern or you may not. So, that, depending on that, kind of how would you, that's I would kind of declare if something is random or not.

3 Colin E [chaos]. Everything is random. Umm, I'm going to disagree with that statement. Umm...ahh, there's...yeah, I trying to find an example where something is not random and umm, I guess any time you make a decision, it's not really random because there's factors that are influencing it or if you, if you can predict something and umm, you're able to observe that, then that's not really random because there's a pattern that you're able to see.

Colin appears to use "pattern" in two different ways in Line 2: to reference a term-based pattern in a sequence and to refer to the amassing of outcomes in clumps. In both cases, Colin wants to get a sufficiently large enough collection of values. Colin's view of looking at a sufficiently large enough sequence to see if there is a pattern is certainly consistent with what I refer to as a first-order random/stochastic model (see Table 2) and lays the groundwork for a productive view of distribution (see Table 7). However, Colin remains focused on using the absence/presence of a term-based pattern to state whether something is random (Line 3). This is most consistent with viewing randomness as a form of sequence complexity (Table 5). Colin continues to refer to this meaning for randomness.

Excerpt 11. Colin's sequence meaning for randomness

1 Colin F [sequence complexity]. A sequence is random when you can't find a pattern, okay. A sequence is random when you can't find a pattern to it like the number π. Okay, ummm, I actually didn't read that before I answered, so I'd like to claim that. Umm...yeah, I would, uhhh, given that you have a large enough sequence, ummm, yeah, I would, [that] would closely fit what I would declare something to be random.

2 Neil — How about the next one?

3 Colin — [Statement G-sequence attribute] A sequence is random when you can't find a pattern, but you can use it to predict something in the long-run. Ooooh, ahhh, is that long-term relative frequency? Ummm...(long pause) you can't find a pattern, but you can use it to predict something in the long run...yeah, let's declare that as random because you...hmmm

4 Neil — What are you mulling over?

5 Colin — Well, if you can't find a pattern, then how can you predict something over the long-term run? So, I'm trying to, how would that relationship would kind of connect. [Neil: okay] So, I'm, the first part "if you can't find a pattern" well, that's, you know kind of fits what I would think be, what declares random. But "you can use it predict something in the long-run" well, then at that point you're able predict future values for that, for example if it is a sequence, umm, so if you can predict future values for that sequence, then is that considered finding a pattern? Umm...so that's kind of what I'm mulling over. (To self) a sequence is random if you can't find a pattern, but you can use it to predict something in the long run.

6 Colin — Mostly I say, it's random. Uhhh, cause if you can't find a pattern then you, you can't, yeah, then next value in that sequence you can't predict what that value will be but you can, on the large scale though, in the long-run, you can predict what the values will gravitate towards. So, I would say that's random.

7 Colin — And then H, none of these match my meaning for random. Ummm, I think that F [sequence complexity] does and after kind of chewing on it, G [sequence attribute] does as well.

8 Neil — You were iffy about C [left-field].

9 Colin — C. Yeah. Umm, I was iffy about that just because, I mean it depends on how many, like if the neutral friend picks a number. Well, what bounds are we putting on that? I mean if they choose from negative infinity to infinity, well, I would consider that as random. But if they say choose, you know one and two, ummm, I don't think that's a large enough sequence to be able to declare if something is random or not. umm...I guess that would be, I guess it is too small of a sequence to declare that it is random. [Neil: Okay] Because in order to declare that something is random you have to have enough values in your sequence to confidently say that there's not a long-term pattern.

Colin endorses the idea that statement F matches his meaning for randomness. His responses leading up to this point indicate that he holds a sequence complexity view of randomness and statement F [sequence complexity] is meant to be interpreted as being indicative of that level in the randomness construct map. However, Statement G [sequence attribute] pushes Colin in that he must wrestle with an explicit lack of a term-based pattern paired with the ability to predict. In line 5, he re-affirms that no pattern in a sequence indicates randomness. However, in line 6, Colin makes a critical jump in his thinking: he brings up predicting a measure of center. While other researchers would view his bringing up what the values "gravitate towards" as being key (e.g., Arnold & Pfannkuch, 2014; Bakker & Gravemeijer, 2004; Reading & Canada, 2011), the more important aspect is that Colin made a distinction between predicting the next term in the sequence and predicting something about the long-run behavior. This distinction provides a fertile ground for Colin to move from viewing randomness as sequence complexity to an attribute. This distinction also highlights that when Colin is typically looking for a pattern in a sequence, he is looking for a term-based pattern. However, he is open to idea that there may be larger patterns that aren't term-based that can be discussed; for example, what value sequence tends to produce.

In line 9 of Excerpt 11, Colin applies his sequence meaning for randomness to the situation involving three friends. However, he confounds the domain of the stochastic variable with the idea of sequence. While the size of the domain does not impact whether a process is random (except, perhaps a size of one), Colin is consistent in his want of large sequences in order to investigate whether or not something is random. To ensure that I understood what he meant by the terms, I asked him to explain what he meant by the words "sequence" and "prediction (Excerpt 12).

Excerpt 12. Colin explains “sequence” and “prediction”

1	Neil	Okay. So, you've used two terms quite a bit [Colin: umm-hmm (agreeing)] in the last conversation. And the first one is "sequence". What do you mean by "sequence"?
2	Colin	A sequence, ummm, so a, in this case it would be, kind of like a...when you have like a set, so it would be the, a sequence can be either numbers...yeah, a sequence would be numbers in a set, ummm, and depending on what the sequence is generated from, for example, uhh, that can be either like an ordered set, it can be, uhh, something that doesn't have an ordinal, nominal, uhh, so I would consider a sequence is just a set of data in that case.
3	Neil	So you said that ahhh, what the sequence is generated from. So, what's generating the sequence?
4	Colin	Ummm, so, for example if you go, well, that depends on, (unintelligible), so that could be the people's answers, so they are the ones generating, so each individual answer when complied would generate the data.
5	Neil	So what would that mean for what a term in a sequence represents?
6	Colin	A term would represent one person's answer. Umm, so I guess each, each sequence, well maybe it's each term in that sequence would be one iteration of whatever uhh, process that you're using to gather the data.
7	Neil	Okay. The other word that you used quite a bit was the word "predict" [Colin: umm-hmmm (agreement)]. What do you mean by "predict"?
8	Colin	Predict. Umm...so this is where...like…if you were, if you were looking at a series and you're to predict what the next value is, what I would, (sigh) umm, if you can make, I guess, predict an educated guess, in that sense, umm you can declare what the next number in that sequence would be, you're predicting what that number would be, you're umm
9	Neil	So if I gave you a sequence and you guess "5" as the next term…
10	Colin	I would be predicting that the next, ahh, term in that sequence is five so
11	Neil	Do you have to be right for that to be a prediction?
12	Colin	No. Cause it's, ahhh, I'm...I'm predicting, so I'm, so whatever process I used to come to that number that's what I'm declaring as being, that's what I'm using to find that number, so if that process is wrong I'm still predicting that number, it's just that I'm not predicting the right number. [Neil: okay] Term or, yeah.

Colin views sequences as the stringing together of multiple trials from some process. Each term in the sequence reflects that datum associated with a person (or object) gotten by carrying out this process. Colin's view of prediction is like that of an educated guess, but he includes the anticipation of not being correct with his guess and that being acceptable.

Colin's conveyed meaning for randomness entails the absence of a term-based pattern in a sufficiently long sequence as well as a separate everyday meaning for the term. Later on, he brought up the two meanings again when he was asked to explain the concept of randomness.

Excerpt 13. Colin explains "Randomness"

1	Neil	How would you explain the idea of "randomness" to another person?
2	Colin	Ahhh, man, ummm
3	Neil	I never claimed these were easy questions.
4	Colin	No, they're not (both laugh). I feel they could be; I feel like it would be easier to explain the idea to a math major then to a random per-oooo-to some person ummm
5	Neil	So why did you go "oooo"?
6	Colin	Well, just because, like when I said "a random person", right, well, that's where there's two kind of uses of "random" that happens in my brain; one is the math and other one is, I guess, some un...I guess, there's, kind of an unconnected, I guess, ahh
7	Neil	A non-math?
8	Colin	Yeah, I guess, well, yeah, a non-math kind of, everyday kind of use that people use, that people, that you use every day in society versus a math kind of actual definition of it [Neil: Okay] Umm…
9	Colin	So how would you explain the idea of randomness to a person? I would try to probably explain the math definition that exists in my head. Ummm, basically I would say...if you, whatever you're observing and you observe this an infinite number of times, if you cannot guess what the, well, if you...if you can't...if you can't guess with some educated guess what, I guess if you can't observe, yeah, if you can't observe a pattern, if you would see this event and then you don't see a pattern occurring within an infinite number that occurs, then, is, that's random.
10	Neil	Okay. How would you explain, ahh, the everyday usage?
11	Colin	Uhh, well, I think people, when people use, at least the everyday usage of "well that's random" what they mean is kind of a shorthand of "that didn't continue with my line of thought" or "that

seems out of sequence" like a non-se, like a non sequitur kind of thing.
12 Neil Okay.

In lines 4 and 6, Colin catches himself using the word “random” in a manner that wasn’t consistent with his “mathematical” meaning for the term. For his mathematical meaning, Colin conveys once again having infinitely many observations to look for some pattern; if there isn’t a pattern, then you have randomness (line 9). Colin re-iterates his everyday meaning for randomness as instances that do not appear to follow the listener’s current flow of thought (line 11).

Excerpt 14. Colin’s View of a Random Variable

1	Neil	What do you think a random variable is?
2	Colin	A stochastic variable.
3	Neil	So what is a stochastic variable?
4	Colin	A random variable. Umm, so a random variable is...a umm...yes...(sighs)
5	Neil	So what are you thinking?
6	Colin	Umm...so you have a variable and I guess as long there's no outside...fac-that's not really...a random variable is...a random variable is a variable that...(long pause) any, so any value or att- any value of an attribute that that variable represents umm, is going, for example if the variable is one out 100, umm, each one of those values has a...equally likely--no, that, that's not, that's not really, if you were to look at that variable and of all, yeah, if you were to look at that variable long-term, like to the infinity, you will (won't?) be able to discern some pattern from that variable.
7	Neil	You won't?
8	Colin	You won't be able to discern, umm, a pattern, yeah, so.
9	Neil	Okay.

Colin’s conveyed mathematical meaning for randomness returns and appears to drive his meaning for a random [stochastic] variable in that if you were to look at the variable in the long-run, you wouldn’t be able to find a pattern (Excerpt 14). He does briefly struggle with whether equi-probability plays a role in a random variable.

Ultimately, he rejects this idea for looking at the variable in the long-term. I suspect that he's imagining a process to get data (such as he described in Excerpt 12) to get form a sequence of values. However, Colin's meaning for a random variable foreshadows the potential for his meaning for randomness to limit his development of the distribution concept. Ultimately, a distribution is a pattern of long-run behavior, albeit a non-term-based pattern. Whether Colin would recognize such a pattern as being able to exist without destroying randomness is unclear.

Danielle's Meanings for Randomness

Danielle initially read through the situations and viewed Situation C (you and two friends) and Statement F (sequence is random when no pattern) as being her top two contenders. She settled on Statement F as being closest to her meaning for randomness. In discussing her answers, Danielle brought up that she felt that her meaning for randomness had evolved over time (line 2 of Excerpt 15).

Excerpt 15. Danielle explains her two meanings for "random"

1	Neil	Okay. So, let's come back to A [left-field]. [Danielle: Okay] So why not A?
2	Danielle	That whole, that whole thing is random. Umm, so, when I think of randomness, umm, if you were to ask me like back in maybe high school [Neil: mmhmm] -ish, umm, to pick something randomness, maybe A would have been a qualifier. But now, especially like as far as I got in college, I tend to go away from the superficial meaning of randomness and towards an actual like quantifiable meaning of randomness.
3	Neil	So what do you mean by a superficial...
4	Danielle	Ummm...useless (laughs)
5	Neil	Let me rephrase: so, what, what do you mean by a superficial meaning for the word "randomness"?
6	Danielle	Oh, okay, ummm, so we're given a context of this, ahh, to Harry Tom's response is random, but to Tom, it isn't. [Neil: Okay] There's a reason Tom said it, there's a connection there somewhere. Umm, so, to say "that's random" in a conversation to me is very

superficial cause it's obvious not random, there's something that had to lead to that point.

7 Neil — Okay. And then you, you juxtaposed that to a quantifiable meaning for randomness. (Danielle laughs) So what do you mean by that?

8 Danielle — Well, I mean, the essence of random isn't exactly, you know, umm, quantifiable, but as opposed to like "that's random" in response to Tom, we can find a pattern by asking Tom, ummm, which means therefore it's not random. Umm, if it was truly randomized, we wouldn't be able to see the pattern or at least not yet.

Danielle's two meanings for "random" share a common link. While her "superficial" meaning focuses on knowing why something happens, her "quantifiable" meaning focuses on knowing a pattern. However, in both meanings there is a strong sense of keeping track of from whose perspective we're making the determination. As Danielle points out, to Tom, his statement of Linda's religion is not superficially random but to Harry, the statement is superficially random. She ties the knowing perspective to her "quantifiable" through the idea of looking for a pattern. This perspective-based view continues to hold in Situation B (someone knocking on your door) as well as Statements D (nothing is random) and E (everything is random):

Excerpt 16. Danielle's response to situations B, D and E

1 Neil — So what about B [unknown/unpredictable].

2 Danielle — Similar reasoning. [Neil: as…?] Umm, someone knocks on your door, err my door, and I don't know who it is, I don't who it is, so the event is probably random for my day, but whoever knocked on my door it's not.

3 Neil — Ahh, and D [ordained], you, you immediately disqualified D. (Danielle laughs) So why's that?

4 Danielle — I've seen like, I mean there's, there are things in this world that I'm assuming like the number π, I don't know if there is a pattern to the number π, umm, there are bound to things that are randomized to different people whether we can explain to them now or if we can explain them later. So, it is not so much that, you know, nothing is ever random, or everything is random, there are things that are random and things that aren't.

5 Neil — Okay. So that would explain E [chaos] as well.

6 Danielle Yeah (laughs)

In line 2, Danielle explicitly draws the matter of perspective into the situation but noting that only for her the event is random but not so for person knocking on her door. She brings this matter up again when discussing why not everything is either random or not (line 4). However, Danielle's two perspective-based meanings appear to fail her when she encounters Statement G (Excerpt 17).

Excerpt 17. Danielle struggles with statement G

1	Neil	Umm, what about G [sequence attribute]?
2	Danielle	That one kind of confused me, ummm, a sequence is random when you can't find a pattern, but you can use it to predict something in the long-run...umm...I don't know, that one's a little harder than the rest of them to answer.
3	Neil	So what do you think that statement is trying to convey?
4	Danielle	Probably, like, π is used to link different formulas...umm, so like, to me π is just a random bunch of numbers strung together but you technically can use it to predict things accurately. [Neil: Okay] Umm, (softly) I'm not really sure about that...
5	Neil	Okay. And you settled on F [sequence complexity] or C [left-field]?
6	Danielle	Ahhh, F, yeah, F. [Neil: F?] Yeah, F.
7	Neil	That was your best match?
8	Danielle	Yeah. [Neil: Okay] C was okay, but there's bound, you're bound to find a pattern with people. So,…
9	Neil	So, like your neutral friend might have a favorite number. [Danielle: yeah] Okay. Alright.

I believe that Danielle's struggle with Statement G [sequence attribute] stems from her meanings. Recall that in Excerpt 15 Danielle described two meanings for the word random: one that she labeled as being a superficial usage of the term such as when two co-workers discuss a third's religion and one she labelled as being more quantifiable—to indicate that not knowing why the event occurred . She cannot apply the superficial meaning to Statement G as she does not know why each term is generated and she takes the statement’s explicit reference to her not being able to find a pattern but still being able to make predictions conflicts with not knowing. This highlights that finding a pattern is the key to making predictions for her. However, she attempts to resolve her struggle by referencing the use of π in mathematical equations. (Recall that Statement F references π as a sequence without a pattern.) Danielle uses her experiences in using π in formulas to get values as the way to understand “predict something in the long-run”.

Danielle brought up the idea of randomness unexpectedly during the portion of the interview focused on the concept of chance. In particular, Danielle appears to view “chance” and “randomness” as referencing the same thing.

Excerpt 18. Danielle explains “Chance” and “Randomness”

1	Danielle	Umm, how would you chance to another person? Chance...(laughs) alright, chance is probably closest to...like randomization. So, in my head, if something were to happen by chance, like the likelihood of it happening would be randomized.
2	Neil	What do you mean by randomized?
3	Danielle	Umm, no discernible pattern to the instance.
4	Neil	Okay. Anything else?
5	Danielle	I don't so [Neil: okay] (laughs)
6	Danielle	(Danielle scoffs, Neil laughs as screen changes) That's why you asked. Alright, umm, how would you explain the idea of randomness to another person? No discernible pattern. (laughs) [Neil: okay] So, whatever happened or whatever's going on, there's no discernible pattern to whatever it is.
7	Neil	Okay. What if there is a pattern but you just don't see the pattern? Would that still be something that's random?
8	Danielle	Yes, until someone can tell me the pattern [Neil: okay]. Until I see the pattern, it would be random. So, random acts probably, could probably have a pattern, but until the pattern is revealed, I would consider it random.
9	Neil	So let's turn back to earlier, umm, somehow I managed to get two pages in between, ahh, we were talking about Tom and Harry [Danielle: Yeah], that match pattern [Danielle: Yeah] umm, you had sort of talked about, ahh, particular with the door knocking, for you that seemed, that would be random but for whoever was at the door that wasn't random. [Danielle: Yeah] So what, ummm, would

		it be accurate to say that for you, randomness deals with who's doing the looking?
10	Danielle	Yeah, that would be pretty accurate. [Neil: Okay] I guess whose ever description of random is...
11	Neil	So that if we had a sequence and I knew, I saw a pattern, but you didn't...
12	Danielle	Yeah, yeah, until you told me the pattern I would consider it random.
13	Neil	Okay.

In Excerpt 18, as Danielle explains chance, she makes use of her perspective-based "quantifiable" meaning for randomness (no discernable pattern). Immediately following the question about explaining "chance" was a question to explain "randomness"; hence, Danielle's reaction in line 6. She gave her "quantifiable" meaning for randomness as the lack of a pattern. I wanted to test her perspective-based meaning by asking her whether her not seeing a pattern would change the status of being random. I referenced Situation B (someone knocking on your door) and as well as sequence. She remains firm in her perspective-based meaning; until she knows the pattern, the sequence is random even if someone else sees the pattern (lines 8-12).

Towards the end of the interview, Danielle faced a task where she was presented with a variety of situations/statements to choose from that matched her meaning for "chance", similar to the one for "randomness". As shown in Excerpt 19, Danielle again presented that her meaning for "chance" was the same as her meaning for "random". Her view that "chance" was the more colloquial term for "random" makes me suspect that if prompted, she should use "chance" as the label for her "superficial" meaning for randomness.

Excerpt 19. Danielle's meaning for "chance"

1	Danielle	Which one of the following most closely matches your meaning for chance. Oh, another one of these. Alright, ummm…
2	Danielle	Chance is the ratio of how many ways you can get a particular outcome compared to the total number of outcomes. (pauses) Chance is the long-run relative of frequency, err, long-run relative frequency of observing some event. Chance is how often you see some event occur in a set of observations. Chance is the likelihood that you have for observing some event. Chance is the probability that you have for observing some event.
3	Danielle	So, chance and random are nestled pretty close together. I would probably use them interchangeably in common, just conversationally.
4	Neil	So chance and random you would use interchangeably [Danielle: Yeah] Okay.
5	Danielle	Umm, if I were, I think turning in an assignment I would use random over chance just because chance is more like a colloquial term, I guess.
6	Neil	Okay. So, would you say that none of these match your meaning?
7	Danielle	Ummm...(long pause) yeah, I don't think any of these would be something that I would choose if I had, if I didn't have to.
8	Neil	I, I won't force you (Danielle laughs) to choose one of them.

Implications

These three students generated five meanings for randomness (Table 9). Both Colin and Danielle came up with two distinct meanings for the "random" while Bonnie made no such distinction. In addition, Colin and Danielle also brought up the notion of looking for a pattern. While Colin's responses give us insight into the fact he was looking for a term-based pattern, Danielle was not as clear, and I did not probe. However, given her responses, I feel confident that she also was looking for a term-based pattern.

Table 9. The Students' Conveyed Meanings for Random.

Student	Meaning for Randomness
Bonnie	A random event is one where the outcome will impact the larger situation.
Colin	A random event is anything that is non sequitur; something that does not fit a person's anticipated flow of a situation.
Colin	A random sequence is one that does not have a term-based pattern when looking at a sufficiently large enough number of terms.
Danielle	A random event is where you don't know why the event happened; something that happens by "chance".
Danielle	A random event is one where you cannot discern a pattern.

Colin's second meaning stands apart from the rest of the student conveyed meanings in that this is the only meaning which makes an explicit reference to needing a sequence with a large number of terms. While Bonnie could imagine many possible outcomes, she judged whether the outcome would impact the larger situation she imagined; Danielle made no reference to needing many terms in order to determine a pattern. These similarities and differences provide for several implications including revising the construct map for randomness and the teaching of randomness.

Revised construct map. The conveyed meanings that the students gave in the clinical interviews do not line up nicely with my initial construct map for randomness (Table 5). However, these students' responses afford me the opportunity to revise the construct map to clarify existing categories and create needed additions (Table 10).

One of the changes I made in the revised construct map is separating out what the student sees as being "unknown"; the outcome or the reason. This distinction affords the opportunity to split out those students who focus more on the why behind the outcome. The student's focus on the reason can serve as a basis for getting the student to start thinking about a generating process as opposed to staying focused on the end result of the process. I made a similar distinction between discerning a pattern and attending to the absence of a long-run pattern. In both cases, the student focuses on searching for a term-based pattern; what Batanero and Serrano (1999) refer to as a regular pattern or Kaplan et al. (2014) view as without order or reason. However, in the former case the student does not express any need or desire to have as many terms as possible in the sequence under study. Additionally, the discernment (or lack thereof) a pattern meaning affords a student the ability to say that the same sequence can simultaneously be random and not random. This incorporates the perspective-dependent element of Danielle's two meanings.

The new Random as Unknown Reason level also shares this aspect. While also focusing on the lack of a term-based pattern, the Absence of Long-run Pattern entails the student wanting to have as many terms as possible to ensure that she cannot find a term-based pattern. This level is distinguished from Sequence Complexity by the fact that in the later the student does not only look for a term-based pattern that could expressed as a mathematical formula. Rather the student is also predisposed to looking for other descriptions that would simplify the sequence.

I added a new level, Randomness as Chance, to capture the view of randomness where randomness is inexplicably tied with imagining outcomes that are equiprobable such as those described by Bennett (1993), Kaplan et al. (2014), and Kuzmak (2016). While Danielle made an explicit reference to "chance" and "randomness" referring to the same concept, she did so in a way that is not consistent with what the previously mentioned researchers found. Rather than making "random" alternate to the idea of Laplacian/Classical probability (a.k.a. "chance"), she made "chance" a subservient to her

Table 10. Revised Construct Map for Randomness.

<table>
<tr><td colspan="2">Attribute of a Process
The student conveys of “randomness” as a property of a process that entails an image of unpredictability in short-run, while anticipating the predictability in the long-run and minimizes sources of bias. A random process will produce a sequence that has 1) no term-based pattern, 2) a sufficiently complex description, and 3) adheres to the Principle of the Impossibility of a Gambling System. Adapted from (Kolmogorov, 2013; Liu & Thompson, 2002; von Mises, 1981).</td></tr>
<tr><td colspan="2">Sequence Complexity
The student conveys that a list/sequence is “random” if the individual’s attempt to describe the list/sequence is to essentially repeat the sequence as given. The individual cannot condense/reduce the list/sequence to a term-based pattern or set of rules that is less complex than the sequence as given. Drawn from (Falk & Konold, 1994).</td></tr>
<tr><td colspan="2">Absence of Long-run Pattern
The student conveys that a sequence is random provided that you have a sufficiently large enough number of trials from the generating process to ensure that there is no term-based pattern to the sequence.</td></tr>
<tr><td colspan="2">Lack of Discernable Pattern
The student conveys that “random events” have a lack of a discernable pattern. Until the pattern becomes clear to the student, she will view the events as random even while acknowledging that to someone who sees that pattern, the events are not random.</td></tr>
<tr><td colspan="2">Randomness as Chance
The student conveys that a “random event” is an outcome that occurs out of a collection of other possible outcomes, each of which is equiprobable. Drawn from (Bennett, 1993; Kaplan et al., 2014; Kuzmak, 2016)</td></tr>
<tr><td colspan="2">Left-field or Non Sequitur
The student conveys that events such as sudden switches in conversation topic, unanticipated question, and unexpected images as being “random”. Inspired by (Liu & Thompson, 2002).</td></tr>
<tr><td colspan="2">Random as Unknown Reason
The student conveys that a “random event” is one that happens but the student does not know why. For example, upon hearing unexpected knocking on a closed door, a student with this way of thinking would say that the event is random to her because she does not know why someone is knocking on the door. However, to the person knocking, the event is not random.</td></tr>
<tr><td colspan="2">Random as Unknown Result
The student conveys that a “random” event is equivalent to not knowing the outcome. For example, upon hearing knocking on a closed door, a student with this way of thinking will say that some “random” person is at the door since he does not who is at the door. Drawn from (Saldanha & Thompson, 2014).</td></tr>
<tr><td colspan="2">Impacting the Situation
The student conveys that a “random” outcome impacts a situation in a way that an alternative outcome would not. The student imagines that situation as continuing on from what is presented and can incorporate affective consequences in her determination of whether the outcome impacts the situation.</td></tr>
<tr><td>Ordained
The student conveys that “random” events result from a chain of events meant to occur. Thus, the student believes that nothing is random.</td><td>Chaos
The student conveys that all events are random and that whatever happens is the result of happenstance.</td></tr>
<tr><td colspan="2">Other
The student’s conveyed meaning for randomness does not fit any of the other levels.</td></tr>
</table>

concept of randomness (a lack of a pattern). Even though none of the students in the present study gave any indication of thinking in this, I felt that I needed to explicitly include this level to help position existing research in the construct map. I must also stress that even though this is the five highest level in my construct map, I believe that there is a great distance between this meaning and the more productive meanings.

An important change I've made the construct map centers on what the student modifies through her usage of the word "random". In the lower levels (Randomness as Chance and down), the student uses "random" to modify the outcome of a process. The Lack of a Discernable Pattern level serves as a transition point to where a sequence of outcomes becomes the "random" object. The sequence remains the object of focus until the highest level where the student gives the underlying generative process the attribute of randomness. My adjustment brings the construct map in tighter alignment with the arguments made by Wagenaar (1991), Falk (1991), and Liu and Thompson (2002) about the need to help students focus on the underlying process. Further, the students in the present study only ever referenced either the outcome or the sequence as being random, not the underlying process. Even when Colin and Danielle brought up some type of underlying mechanism, they did not go so far as to apply the label of "random".

Teaching randomness. There are several implications to the teaching of randomness that bear discussion at this point. Perhaps the most important implication this study has for teaching is the clarion call that we as educators pay careful attention to our language. As Kaplan et al. (2009) found and was seen here, students are aware of "random" being used in multiple ways. Making students aware of technical and non-technical usages is important to get students to build a coherent and productive meaning for randomness that can then be leveraged by later concepts in Statistics. However, this means that educators must be on constant guard to ensure that they do not use a non-technical usage in front of students or in course materials. As Hatfield (2016a) noted, a teacher's conveyed meanings can act as a limiting factor for students' meaning development. This also places an impetus on educators and researchers to critically examine what meanings for randomness are being taught. The top-tier meanings espoused by the extant literature (e.g., Kaplan et al., 2014; Kuzmak, 2016) need to give way to those meanings which are much more productive (e.g., Falk, 1991; Falk & Konold, 1994; Kolmogorov, 2013; Liu & Thompson, 2002; Saldanha, 2016; von Mises, 1981; Wagenaar, 1991).

Further, instructors need to take steps to problematize those meanings that are in middling levels of Table 10; in particular, levels Random as Unknown Result through Absence of Long-run Pattern. These meanings function as useful in the moment meanings for students who hold them. Students who have constructed these meaning are rarely, if ever, challenged in curricula to wrestle with identifying whether or not something is random, let alone judging a process to be random. Most curricula introduce randomness as a necessary tool and then cover randomization procedures. In homework, students are at most asked to describe and/or carry out a randomization procedure (i.e., using a random number table). Carrying out such a procedure does not require a student to have to face the implications of her meaning for randomness. Rather, we need to ask students to describe what is meant by the term "random" and ask students to classify whether or not different processes are random. The Sequences applet I've built provides a fertile site for students to repeatedly engage in activities to help them productive meanings. Other questions that students could wrestle with could include presenting students with a sequence and asking them to explain why they believe the sequence came from a stochastic/random process (or not).

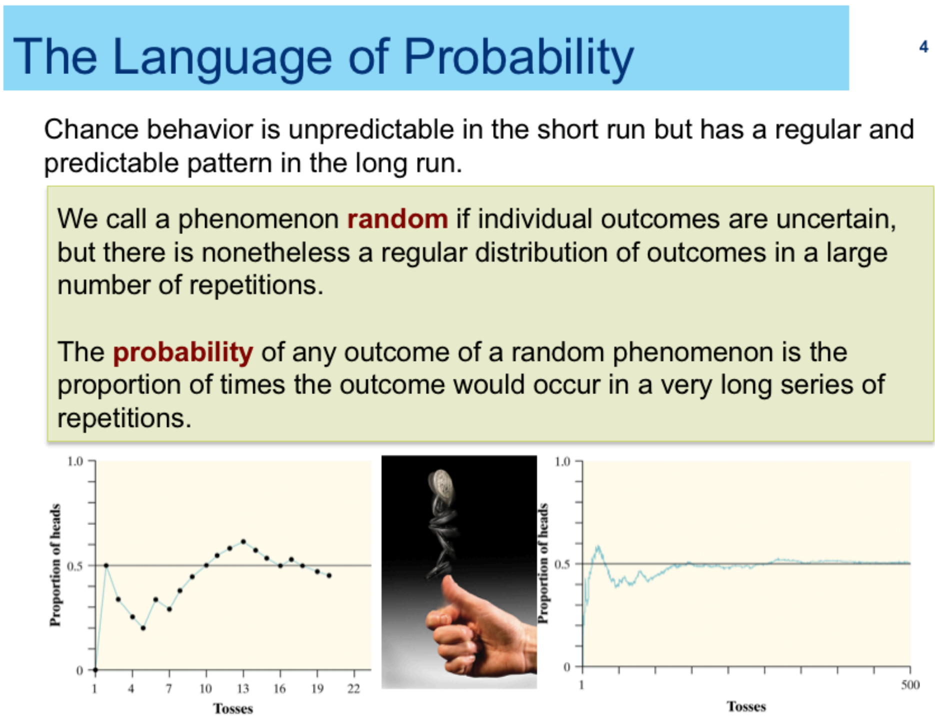


Figure 23. Slide introducing randomness (D. S. Moore, McCabe, & Craig, 2017b)

Such activities do hold potential to help students develop higher order meanings in the construct map. Recall that Colin's introductory Statistics course for bioscience majors was a reformed, conceptual approach. Within that course, there were multiple discussions on the various meanings for randomness as well as question like those described above. Partnered with the discussions on randomness, there was also an emphasis on looking at the long-run when discussing ideas of distribution and moments. While Colin did not convey a meaning consistent with the meanings espoused by the course (Attribute of a Process and Sequence Complexity), his conveyed meanings still carry hallmarks of that course; a focus on sequences and the long-run (i.e., infinitely many trials). His fall back to searching for a term-based pattern could be indicative of the traditional Statistics course he was enrolled in during the study. Figure 23 shows a publisher made slide meant to accompany the ninth and newest edition of Moore, McCabe, and Craig's (2017b) *Introduction to the Practice of Statistics* text. This serves an example of what most traditional introductory courses present randomness as randomness; Colin was working out of the eighth edition. Such a slide presents randomness in such a way that students can continue to use their useful-in-the-moment, unproductive meanings without problem. Notice that the slide even sets students up for the same chance-as-randomness meaning that Danielle gave; "chance behavior" becomes "random phenomenon". While the given meaning has hallmarks of the meaning that Liu and Thompson (2002) espouse for a stochastic process, D. S. Moore et al.'s presented meaning maintains a focus outcomes rather the underlying process. As a reminder, the meaning needs to be read with the notion of distribution as the thing that tells us what values a stochastic variables takes and how often these values are taken. Should an instructor follow the course flow recommended by the authors, students will be seeing this slide after already equating "randomness" with "no pattern" (Figure 24). Such curricular materials could limit students to at best reaching the Absence of Long-run Pattern level of the construct map (Table 10).

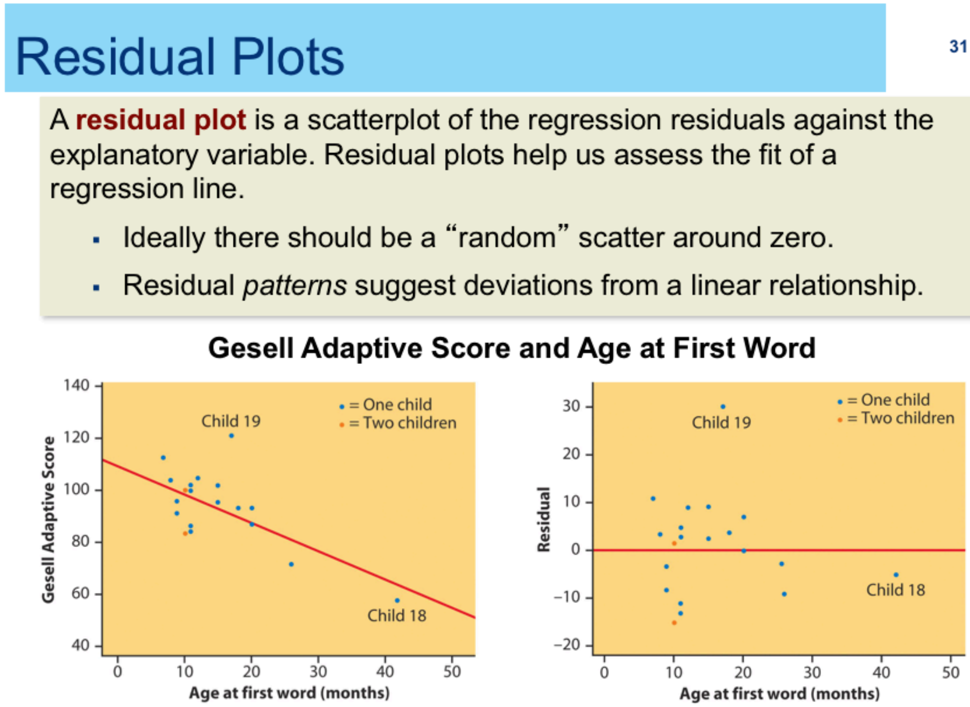


Figure 24. Slide tying randomness to "no pattern" (D. S. Moore, McCabe, & Craig, 2017a)

The meanings for randomness that Bonnie, Colin, and Danielle gave in addition to the published research point out that there is still a need to for researchers and educators to critically reflect upon the idea of randomness. Statistics education researchers need to focus in on the meaning for randomness that we want to support students in developing in relation to their development of other productive and coherent meanings in Statistics. I believe that helping students develop a meaning for randomness that pins randomness as an attribute of a process is the most coherent conceptualization of this idea. Building a curriculum that supports students in building this meaning presents an opportunity for future research.

Chapter 7: Exploring Bonnie's Meaning for Stochastic Processes

After her initial clinical interview, Bonnie came in for two additional sessions in the vein of exploratory teaching interviews. In the first session, she worked through the Light Switch (Version A) and the More Processes Activities (see Chapter 5: Methodology). The second session centered on the Sequences Activity. The intent was to see what sense of stochastic processes she would develop through interacting with these applets.

Bonnie's First Session

At the onset of the first exploratory teaching interview, I asked Bonnie if she had ever heard of the phrase "stochastic process" before and when she responded no, I asked her if she had heard the phrase "random process" before; has her reply:

Excerpt 20. Bonnie's initial explanation of random processes

1	Bonnie	Yes.
2	Neil	So where have you heard that phrase?
3	Bonnie	Umm, I've heard it in Statistics, in my class last year. Umm, during, I'm not sure exactly which unit, but I remember hearing it sometime during that semester.
4	Neil	Do, uhh, so what is a random process?
5	Bonnie	Umm, I think this relates back to random variables, or randomness, what we were talking about in our last session. [Neil: okay] I think it could mean, umm, all the variables in a situation and the outcome are all determined by randomness.

While Bonnie hadn't heard of a stochastic process, she did have exposure to the phrase random process, but her meanings for this phrase are vague. Her meaning for random process appears to be driven entirely by the term "random" which she links back to both random variable and randomness in the clinical interview. During the clinical interview, Bonnie had explained a random variable as being the outcome of a process that will affect the individual(s) involved in the imagined situation (Excerpt 21). Bonnie's meaning for randomness is that the outcome of the process has to create some kind of effect in the situation; for example, being nervous about speaking (Line 6).

Excerpt 21. Bonnie's meaning for random variables

1	Neil	What do you think a random variable is? Do you recall hearing this phrase in Statistics class?
2	Bonnie	Oh yeah. Umm, I believe, so an example I'm thinking about right now is if you're picking out of a hat, let's say, but these variables won't have any affect on the outcome. [Neil: Okay] If they, and...umm...so say there's a word problem and ummm, (long pause) hmmm, (long pause),I just think it is something that does not...err, (sigh)
3	Neil	Could you give me an example of a random variable?
4	Bonnie	Hmmm, (long pause), I'm trying to think right now, and I can't
5	Neil	It's okay; remember there are no wrong answers
6	Bonnie	Yeah. So earlier I defined random as something that most of the time will always have an effect on the outcome...so if you picked out of a hat and...you picked either one, two, three, or four, or five and whatever number you got was the order you went to present. Well, if I got a three that's a random variable but its gonna have an effect on how I'm gonna, how nervous I'll be to present or how calm you'll be to see [Neil: Okay] So, yeah, I think that's my answer.

Bonnie's meaning for random process does not afford her a connection to data as shown in Excerpt 22. For her, data appear and become fodder for data visualizations. This is unsurprising given that many introductory textbooks do not place much emphasis on stochastic/random processes but jump to giving students data with which to make graphs and do calculations on.

Excerpt 22. Bonnie's view of the origin of data

1	Neil	Where do you think data come from?
2	Bonnie	I think…it comes from numbers of different categories and they bunch it all together in specific graphs or charts and it's organized in a very specific order.

As Bonnie initially worked with Room One (deterministic, white and black), she quickly assimilated what was happening to her past experiences for turning the light in a room. When asked to create a new switch/button, Bonnie struggled to understand the prompt. However, once she understood that she needed to come up with the details behind the switch/button, she drew a switch and wrote her first rule in Figure 25. Starting from a black room, Bonnie noted that pressing the button once would turn on the light. Her next rule was that double tapping the button would make the light go off (crossed out in the middle of Figure 25). I asked her if we had needed to double tap the switch in the applet to make the light turn off and she corrected herself to say that every other tap would reverse the state of the room (third line). As she went to write the third rule, she hesitated as she tried to come up with the wording. In getting her to explain her thinking, I asked her if we reverse on every tap or every other tap. She initially says every other tap, but in seeing the switch for Room One get pressed again (going from a white room to a black room), she amends her statement to every tap reversing the room.

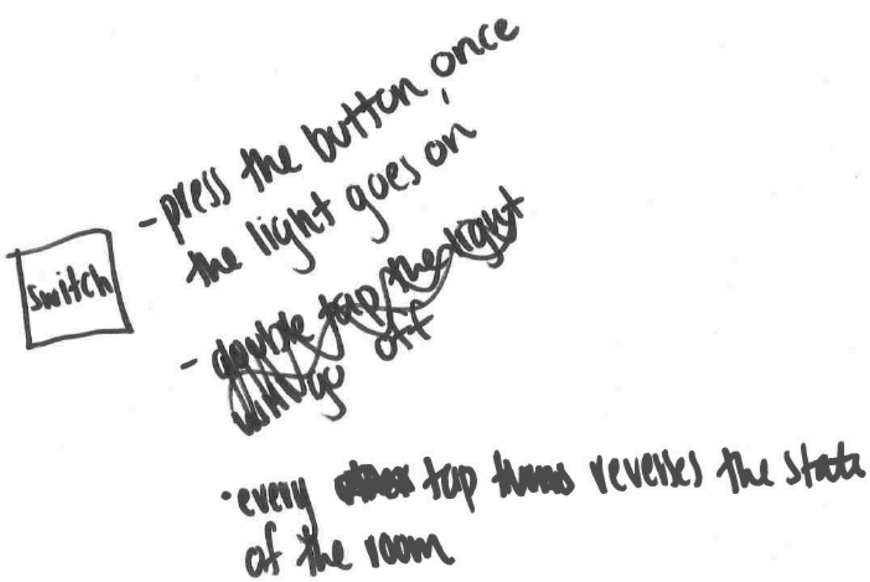


Figure 25. Bonnie's new button for room one

In Room Two (deterministic, grey scale), Bonnie made an unprompted move to draw out the various states of Room Two; she created the cycle diagram shown in Figure 26. Her original drawing had two separate pieces: the upper set of boxes for starting from black and going to white and the lower set of boxes for the reverse. Bonnie quickly realized that the boxes on the outer edges would be the same state (black on the left, white on the right). She proceeded to use her diagram to tackle how to get back to a previous state of the room.

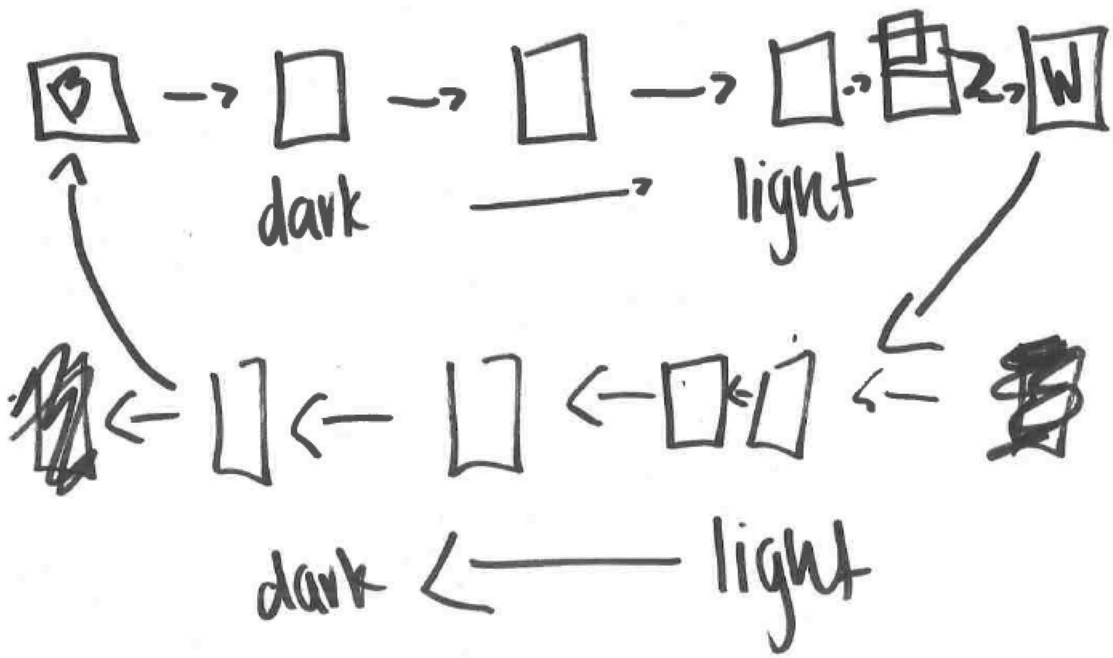


Figure 26. Bonnie's cycle diagram for room two

When asked how she would create a new switch for Room Two, Bonnie had this to say:

Excerpt 23. Bonnie's new button for room two

1	Neil	Suppose that the switch broke [Bonnie: mmm-hmm] How would you, sort of program a new switch, based upon all that you've now discovered about Room Two?
2	Bonnie	Hmmm...I would...(long pause)...it would the same concept that it already has since there's five different stages [Neil: hmm-mmm] or

states of the room, so by every click it will give you a different state for the room [Neil: okay] and then...vice versa so when you hit all five clicks [misspeak? states] it, and you want to back to the darkness, you have to it [the button] once again.

3 Neil — Just once? So, we got all of the way to white, we just hit once to get back to black?

4 Bonnie — No. Each of, you have to hit it five times.

5 Neil — To get back to black [Bonnie: Yes] So each time you press the button you sort of advance one through the cycle. [Bonnie: yeah]. So, we could write that as sort of a process, right?

6 Bonnie — Umm-hmm (Agreement) So, [writes "click switch 5 times, but once for each new state"] So I wrote click switch 5 times, but once for each new state.

7 Neil — And what would clicking the switch 5 times accomplish?

8 Bonnie — Umm, going from complete darkness to complete whiteness. [Neil: Okay] But by hitting it once, that goes to each different state [Neil: Okay]. So, rather than saying "you only hit it 5 times", well you, hit it five times, you're not gonna just go from the next state, you're going to go all the way to whiteness [Neil: Okay.]

Bonnie's cycle diagram afforded her the ability to think not only see the different colorings of the room but also gave her a way to see the action of pressing the button/switch as moving her forward through the cycle (Line 2). This allows her to more readily address her initial error of one click back to black (Lines 2-3). While Bonnie embedded the notion of clicking 5 times in her new switch's rule (line 6 of Excerpt 23 and Figure 27), I believe that she did so for two reasons: she might have been focusing on the idea of going from a black (completely dark) room to a white (completely lit) room and/or she might have been trying to indicate that there were four intermediate states in order to get to the desired state of the lit (white) room. Her response in Line 8 supports this notion.

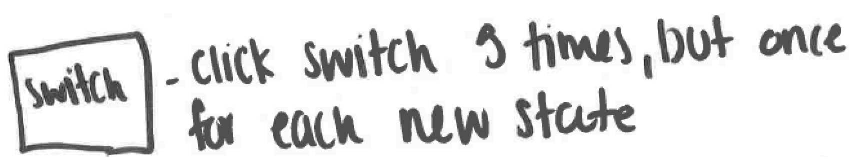


Figure 27. Bonnie's new switch for room two

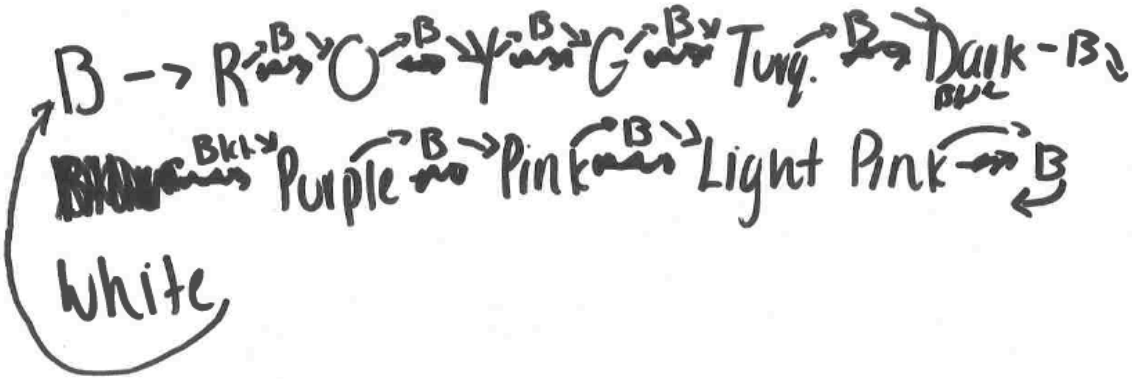


Figure 28. Bonnie's cycle for room three

Room Three (deterministic, rainbow) introduced added complexity through the presence of colors outside of the typical experience of turning on the light in a room. As Bonnie played around with Room Three, she recognized that the light in the room was moving through the color spectrum. Unlike for Room Two, she did not write out a cycle; however, she was able to generate one when prompted (Excerpt 24).

Excerpt 24. Bonnie's building a cycle for room three

1	Neil	Do you think you could come up with a, uhh, a cycle pattern like you did in Room Two? [Bonnie: yes] Do you remember what color we started with?
2	Bonnie	Yes. Do you want me, like to write the first letters?
3	Neil	Sure.
4	Bonnie	So we started with black, (softer) went to red, then orange, yellow, green, let's see, a turquoise, and then...[stops talking, continues to write colors]
5	Neil	So, okay, how could you check your sequence?
6	Bonnie	So, I'll start with black [clicks]...wait! What?!
7	Neil	So you're expressing surprise. [Bonnie: laughs]. What, why are you surprised? [Bonnie: yes]
8	Bonnie	I didn't know that there's another black thrown in there.
9	Neil	Well, let's try. So here we are at black [Bonnie clicks; Bonnie: so then red] And then
10	Bonnie	Black, oh, okay.
11	Neil	So what are you noticing?

12 Bonnie — After each color, it goes back to the state of darkness [Neil: okay] or black. So then between each color there's a blackness, black state.
13 Neil — Okay, so do you want to tweak your...
14 Bonnie — [Bonnie adds black states between her colors]

She initially focused on the colors that showed up after the initial black of Room Three (Line 4). She encountered a surprise when she went to check her cycle against the applet: there is an instance of a completely dark room between each colored light. When she started her check, Room Three was light pink and when she advanced the button, she got a black. In her original cycle, that put her at the start (upper left of Figure 28) so when she clicked next she got white instead of the red she was anticipating. This led her to revise her cycle adding in black states. As she did in Room Two, Bonnie views her cycle diagram as a way to program a new button (Excerpt 25).

Excerpt 25. Bonnie views the cycle as the button program for room three

1 Neil — So, do you have the whole sequence worked out? [Bonnie: umm-hmm (agreement)] Do you think that you could use that sequence as sort of the program behind the switch?
2 Bonnie — Yeah.

Room Four (stochastic) presents an interesting situation for Bonnie. She initially guesses that the room will stay black when she presses the button. When she presses the button the first time, the color of Room Four changes to dark green color. She then guesses that the color will become easier to see if she presses again which results in a burgundy color. Bonnie then anticipates that the color will change but we'll be able to see the color better (becomes grey). She keeps guessing colors and clicking the button/switch to Room Four. After her eighth click in Room Four, Bonnie proposes a pattern:

Excerpt 26. Bonnie's first attempt at a pattern for room four

1	Neil	So what are you thinking about for Room Four?
2	Bonnie	Ummm, so when you started with black and you hit or you tap the switch button, it goes to a color and tap again it goes to another and then it goes to grey.
3	Neil	Okay.
4	Bonnie	And then after that, you're going to go through two more colors and you're going to hit grey again. And then two more colors and then grey again.
5	Neil	So, should the next one be grey or another color?
6	Bonnie	Umm, another color.
7	Neil	Alright. And so, the next one should be...
8	Bonnie	Grey. [Clicks] (softly) Oh no...
9	Neil	Why did you say "Oh no"?
10	Bonnie	Uhhh, I thought it was going to be grey, cause I, I thought that I could see a pattern occurring [Neil: okay] but now, it when through three colors, so maybe it will grey now. [Clicks] No. Okay. I don't, I don't know...
11	Neil	You don't know what?
12	Bonnie	If...there's a sequence or a specific pattern.
13	Neil	So, you're thinking about trying to come up with things like what you did from Rooms One through Three [Bonnie: yeah]. Having problems with this one? [Bonnie: umm-hmm (agreement)]

The applet was designed to give a grey color tone as the fourth and seventh states. In her quest to find a pattern, she does appear to want to come up with a cycle diagram as she was able to in the previous three rooms. Bonnie continues to play with Room Four, expressing surprise with each time that what she guesses does not hold up with the press of the button/switch. At that point in time, we have the conversation in Excerpt 27.

Excerpt 27. Bonnie introduces "randomness" for room four

1	Neil	So what, what are you feeling right now with Room Four?
2	Bonnie	Confused.
3	Neil	Confused about what?
4	Bonnie	Umm, if there is a certain or a specific pattern or not. [Neil: okay] It just seems...almost like it is just random.
5	Neil	What's random?
6	Bonnie	The...order the of the colors?

7	Neil	The order of the colors? [Bonnie: and the pattern] The pattern. Would that be the order?
8	Bonnie	Ummm, no. [Neil: no] I think just the pattern is all.
9	Neil	Do you think there is a pattern?
10	Bonnie	Ummm...maybe.
11	Neil	Maybe? What would you need in order to find out whether or not there's a pattern?
12	Bonnie	I would have how we started with black, complete darkness, I'd have to go through one whole cycle and end up back there and then when I click again, try to remember if those were the same colors that came after
13	Neil	So, could we, could you do something like what you did with Rooms One through Three? [touches her cycle diagrams]
14	Bonnie	Yes. Even though I don't know if it is for sure a pattern.
15	Neil	So let me, I'll reset [Bonnie: okay] So we're at black for Room Four.
16	Bonnie	Okay. [writes B] So, last time it was [writes "dark"]
17	Neil	So how about we just build a new cycle? [Bonnie: okay] So we've got black and then...if you press the switch, what do we get?
18	Bonnie	A new color. So, I think it is random.
19	Neil	What's random?
20	Bonnie	The pattern...or there is no pattern!
21	Neil	There is no pattern? What if I told that there was a pattern?
22	Bonnie	Ummm, I don't think it follows any specific rules [Neil: umm-hmm]
23	Neil	Do you think you could create....let's say the switch broke, [Bonnie: okay] do you think that you could program a new switch?
24	Bonnie	With a specific rule for the room...
25	Neil	So that Room Four functions exactly as Room Four is functioning
26	Bonnie	Okay. So, you would just have your switch button and you would start at black and every time you hit the switch, it would just go to a random color.
27	Neil	Just a color. [Bonnie: yeah] No, no, no finer grain details than that? Like what you had in Rooms One through Three?
28	Bonnie	I think the first, after you hit the first switch, you'll get two random colors and then you'll to a grey stage. [Neil: okay] And then
29	Neil	So, let's test that. So, you have one, your first color. [Bonnie clicks] Second color [Bonnie clicks again]
30	Bonnie	So, no. I don't it follows any specific pattern with rules. [Neil: okay] It's just kind of thrown out there.

Bonnie introduces the notion of randomness in Line 4. The way that she is using the term random here is unlike how she used the term in the clinical interview. There she used the notation of randomness to indicate that the outcome of the described processes would have an effect on individuals in the given situation; here she uses random to mean without a pattern (Lines 5-10, 18-20). When I suggest that there might be pattern, she states that there would not be any specific rules (Line 22). She carries this through with how she would program a new button for Room Four. We can see here that she brought back her two colors then grey pattern. However, when that pattern breaks, she goes back to the lack of rules and no pattern (Lines 26-30).

I asked Bonnie to create two categories for the four rooms and state which rooms were in which categories. She immediately made a Pattern category with Rooms One through Three and a Random category with Room Four (Figure 29).

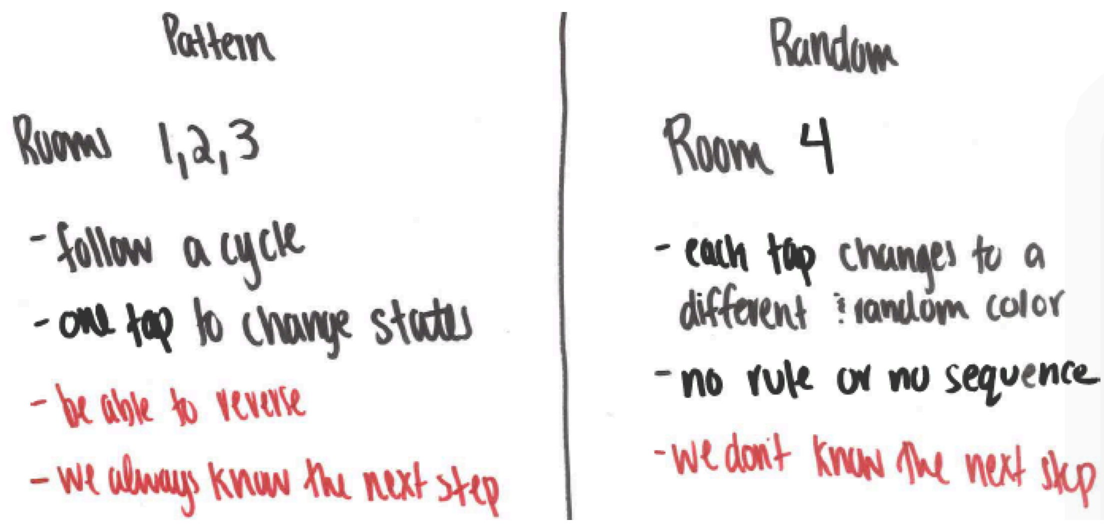


Figure 29. Bonnie's two categories

Her initial descriptions of the categories (in black) match the experience she had in creating her cycle diagrams for the first three rooms for the Pattern group. For the Random group, she wrote the first bullet point and I asked her what she meant by a random color, she wrote the "no rule or no sequence" line.

Having established her categories, I moved into the next activity (More Processes) where I presented here with five new processes and asked her to then use her two categories on five new processes. The five new processes included:

- Mixing fixed amounts of baking soda and vinegar together at room temperature (deterministic)
- A teacher running a lottery with her students' names on Ping Pong balls (stochastic)
- Setting up a savings account at Barclays with 1% APY, an initial deposit of $1000 and no additional deposits or withdrawals (deterministic)
- Looking at the color of Neil's shoes and recording the most prominent color (deterministic; informed that I only wore black shoes)
- Standing at an intersection every Monday, 2pm to 4pm, and counting the total number of red vehicles that went through the intersection (stochastic)

Bonnie's categorizations of each of these new processes appear in Figure 30.

Mixing baking soda and vinegar	pattern ~~random~~ pattern
Ping Pong ball lottery	Random
Savings account	Pattern
The color of Neil's shoes	Pattern
Total number of red cars through an intersection	Random

Figure 30. Bonnie's categorizations of five more processes

The first new process (baking soda and vinegar) was the most challenging of them as evidenced by her changing her classification. Partly this was due to the fact that she did not have the experience of mixing baking soda and vinegar together. She initially states that this process belongs in the Pattern category. When I asked why, we had the following exchange:

Excerpt 28. Bonnie revises her process categories

1	Neil	Why do you think that this is going to all into the pattern category?
2	Bonnie	Ummm...just, ummm, could you read the answers again?
3	Neil	So when you mix these two amounts together, you're going to get 4 grams of water, [Bonnie: okay] 9.77 grams of carbon dioxide, and 36.43 grams of sodium acetate.
4	Bonnie	Okay, so, ummm, I think that the more you have of both of these, the more that you're gonna create, and then the less you have of these, the less you're going to create, which kind of follows a specific pattern [Neil: hmm-mmm]. Umm...but it doesn't follow...actually I'll just stop there.
5	Neil	So why did you, what were you going to say?
6	Bonnie	That there's no, really specific number to add to or multiply to like how in the patterns (category) here we had by one tap you could get to a different state. And for example, in Room Three, we figured out "oh well from turquoise to go back to the black (just prior) it would take 19 taps to go back, which I don't think you could really identify a specific number.
7	Neil	Once we combine these two, can we uncombine? [Bonnie: no] Not really. [Bonnie: yeah] So, sort of like pattern [Bonnie: mm-hmm (agreement), but maybe we have to come back and adjust the hallmarks of a Pattern?
8	Bonnie	Yes.
9	Neil	I'm going ask you to use a different color.
10	Bonnie	Okay.
11	Neil	So, you had this like bright idea, light bulb moment, [Bonnie: yes] on your face when I asked you to change color. What were you thinking?
12	Bonnie	Another factor that plays a role in pattern would be that you have to be able to reverse.
13	Neil	You have to be able to reverse? [Bonnie: Yes] Okay. So, if we add that, will this be a pattern anymore? [Bonnie: no] So would this be random then?
14	Bonnie	Yes. I think so, Yes. [Updates her category description]

15	Neil	So here in your random category, you had each tap changed different color, that was sort of specific to Room Four [Bonnie: umm-hmm (agreement) Yeah] But down here you have no rule or no sequence. What if I told that no matter how times you do this, as long as you add 20 g of baking soda to 30.2 grams of vinegar, you'll always get 4 grams of water, 9.77 grams of carbon dioxide, and the 36.43 grams of sodium acetate?
16	Bonnie	So then it would be...oh...
17	Neil	So would there be a rule?
18	Bonnie	Yes.
19	Neil	Okay. So, would that now fall in random?
20	Bonnie	(pauses) No. I...I...now I think it's pattern.
21	Neil	You think this is pattern again? [Bonnie: yeah] So why are you thinking this is pattern now?
22	Bonnie	From the explanation that you just gave...
23	Neil	That as long as you mix those two amounts you'll always going to get the same end products
24	Bonnie	Yes. Because if you look back to Room Four, uhh, whenever start with starting product, black, well you didn't always get turquoise, you gotten dark green, or pink rather than in these ones [gestures to Rooms One through Three], you always started again at the same spot and you would always move on to the same state, states. Like here, [points to Room Three] you'd hit the switch and it would go to red and it would always go to red during every cycle. [Neil: okay] So, I think it is pattern now.
25	Neil	You think this is pattern. So, let's go ahead and cross that out. What tweaks might you need to make to here [points to her category descriptions]?
26	Bonnie	Umm, I'm not sure how to word it
27	Neil	So, talk me through what you're thinking
28	Bonnie	So, kind of just how I explained to you; you're, you always start at the same base, kind of, even for all four of them, but then for Rooms One, Two, and Three, you're going to continue on the same, to the next level
29	Neil	So, from the same staring point [Bonnie: uh-huh (agreement)] we always know the next. [Bonnie: Yes.] And is that true in Room Four? [Bonnie: No] So how about we just write that?
30	Bonnie	So we always know the next step
31	Neil	And then would that work here then? Given that we start with these things being combined together? We know...[Bonnie: umm-hmmm (agreement)] Okay.
32	Bonnie	So then this would be pattern. [Starts writing]
33	Neil	So what are you adding to random?
34	Bonnie	That we don't know the next step. [Neil: okay]

Bonnie first views the baking soda and vinegar process as an example of her pattern category, but as she explains why she believes so, she runs up against her inability to move beyond her intuitions about more reactants meaning more products (Line 4) to see a clear and specific way to get the end amounts (Line 6). In the moment, I had asked her about reversibility and while she could use her cycle diagrams to get each of the first three rooms back to a previous state, she could not imagine how to undo the experiment. This sparked her to add on the notion of reversing to her description of what goes into her Pattern category (addition of the first line of red text on the left in Figure 29). In Line 15, I mention that if we combine the two stated amounts together, we'll always get the same end result. This causes Bonnie to re-think placing this process in her Random category. What she ends up doing to noting that in her Pattern category we can state what the next time through the process will result in (Lines 24-30). She extends this to add a new feature to her Random category that we do not know the next step (Line 34). Her use of the phrase "next step" refers to the next time we click the button or mix baking soda and vinegar together. At this point in time, her Random category encompasses the lack of a [clear] rule or sequence (i.e., a term-based pattern) but also the inability to state what we'll get on the next time through the process.

Excerpt 29. Bonnie's view of the ping pong ball lottery

1	Bonnie	I think that the teacher, when she picks a ball, it is just going to be random
2	Neil	What's random?
3	Bonnie	The...uhh, the process of it.
4	Neil	Ahh, so the process of picking...the ball is random. [Bonnie: Yes] So does that make this like your pattern category or your random category?
5	Bonnie	Ummmm, well, at first you could say its pattern because we know we're going to pick out a Ping Pong ball from the bag and there's

going to be a name on it. But we don't know exactly what that name is going to be, or who it is going to be. It could be a boy or a girl. [Neil: Okay] So, it's something we don't know the next step of, so I think it would be random [Neil: Okay]

6 Neil [Bonnie writes "random"] We know that, at least prior to the next step that she will have a ball [Bonnie: yeah] right, but we don't know anything about which ball [Bonnie: mmm-hmm (agreement)]. Okay. But, we could, could we imagine her doing this again and again?

7 Bonnie Yes. [Neil: So, like every day she runs this lottery?] Yes.

8 Neil Would we be able to repeat this process?

9 Bonnie Yes, but not entirely the exact the same.

10 Neil So what do you mean by that?

11 Bonnie So, say she picks out, I don't know, a name "Ben" and then the next day she picks out another one, well that could, that name could be "Jennifer", so it's going to different. I mean it could be [the same] but it would be by chance. [Neil: okay] So I don't think that there's a pattern.

The next four processes fairly straight forward for Bonnie. When discussing the Ping Pong ball lottery, Bonnie uses her addition of knowing or not knowing the next step to help decide in which category this process belongs (Line 5 of Excerpt 29). Here she also alludes to the notion of replication (repeating the process but not necessarily getting the same outcomes) in Line 11. While the image of repeating the process is present throughout much of her discussions, she does not make the repetition of the outcomes an explicit part of her categories. Stochastic processes have replication (repeatable process but the outcomes don't occur in the same way) while deterministic processes have reproducibility (repeatable process and the same outcomes occur in the same way). Bonnie classifies the savings account and the color of Neil's shoes as being patterns, citing that we could use a formula for the former and that Neil's shoes are all have black as the predominate color as her justifications. In Excerpt 30, Bonnie discusses why she thinks that the process of recording the total number of reds going through an intersection belongs in her Random category. She refers to the lack of patterns and rules as well as not knowing the next step (Lines 2, 4, and 12). In Line 12, we can once again see Bonnie make reference to the idea of replication that we can repeat the process of going to the intersection and counting, but that the total number of red vehicles might not be the same from week to week.

Excerpt 30. Bonnie's view of the red cars through the intersection

1	Neil	Why do you think that this matches your random category?
2	Bonnie	Umm, I think that it matches it because there's no sequence and rule for when for color cars go on which streets, [Neil: umm-hmm] and again, with the shoe example, I'm not going to turn around and look again and it's going to be the same red car, it could be a blue car or a green car. And, we're not going to know, like the next step of this.
3	Neil	So what would be the next step?
4	Bonnie	Umm, of what color car we're going to see next [Neil: okay] It could be a red one, could be a blue one.
5	Neil	So what are we trying to record [points to the process text] with this process?
6	Bonnie	The total number of red vehicles that go through the intersection.
7	Neil	So, if we imagine that as one iteration [Bonnie: umm-hmm] like one button click. [Bonnie: yeah] The next button click would be...
8	Bonnie	Just random, like Room Four.
9	Neil	Like Room Four. So, it would, we'd be looking at the next Monday's [Bonnie: Yes] Even though we're still looking at Mondays in the same time spot, will we always have the same total number of red cars go through that intersection?
10	Bonnie	Ummm, no.
11	Neil	No. Why not?
12	Bonnie	Ummm, just because it is the same time and same area doesn't necessarily mean that the same numbers are going to go through again. I think it, there's no rule or sequence [Neil: okay] So, someone that drove through yesterday at that time, might not drive through next Monday. Because they might be out of town or they might take a different [Neil: Okay] street. I just think it would be random.

At this point in time Bonnie's thinking appears to have found solid footing. For her, a process belongs in the Pattern category when that process has a clear rule and more importantly, that we can determine what will come next. A process belongs in her Random category when there is no clear rule or pattern and we can't state what will happen next. I end the session by introducing Bonnie to the formal terms random/stochastic process and deterministic process (Excerpt 31).

Excerpt 31. Bonnie learns the names of deterministic and stochastic

1	Neil	There are formal names for these categories. Any ideas what they might be?
2	Bonnie	One of them will be random processing
3	Neil	Or stochastic [Bonnie: Or stochastic]. Okay, stochastic processes and random processes, same idea. Do you know what this one, this Pattern one is called? [Bonnie: Ummm, nnnn] Deterministic [Bonnie: okay]. Why would this one be called "deterministic"?
4	Bonnie	Umm, I can see why because of the root word, you're able to determine what's going to happen next.
5	Neil	So what do you that might mean for the word random?
6	Bonnie	You're not able to determine what's going to happen next. You're unsure.
7	Neil	So we might not be able to determine what happens next. Does that mean that we can't investigate what might be happening over, let's say a decade for the Rural Road and Apache Boulevard?
8	Bonnie	Umm-hmm, no, you can still investigate, you just...
9	Neil	have to take the long view...
10	Bonnie	Yes; and then use that as your data.
11	Neil	So where do data come from?
12	Bonnie	Umm, random processes [Neil: very good] Wow! (laughs) That's crazy!
13	Neil	I feel like your mind has exploded. [Bonnie: yeah it is] So, why, why do think that is crazy?
14	Bonnie	Umm, to me personally, I like to link things together [Neil: mmm-hmm] like in any situation, and when you take a step back and look at the overall picture to see it all connects, I think it's fascinating. (laughs)
15	Neil	Had you ever thought about where data come from before?
16	Bonnie	Umm, I mean, I just knew, you would go out, not out but collect data and actually, you do kind a use random numbers
17	Neil	Wouldn't this [points to red cars through intersection text] be going out and collecting data? [Bonnie: yes] So notice, every experiment [Bonnie: mmm-hmmm (agreement)], you design a process.
18	Bonnie	Yes. Yes.

Bonnie makes a connection between the root of "deterministic" and her pattern category and her primary hallmark for that category: we know what will happen next (Line 4 of Excerpt 31). She then extends this to her random category by positioning random processes as the opposite of deterministic processes. Out of curiosity, I asked Bonnie we could still investigate what might be happening over an extended period of time (moving towards the notion of long-run behavior) even when we couldn't say what would happen next (Line 7). Bonnie replies that you can and brings up data. The light bulbs began to flash for her as she made a connection that she hadn't really thought of before: that data come from random/stochastic processes. This is quite different from her original answer that conveyed that data just appear (Excerpt 22).

As a whole, three key things occurred in Bonnie's thinking. First, Bonnie's meaning for randomness shifted away from the idea that the outcome of the process had to effect actors in the imagined situation (i.e., no more discussion of how the individuals feel as a result of the outcome). Rather, her meaning for randomness shifted towards that of Lack of Discernable Pattern (Table 10). Second, Bonnie's image of stochastic process entailed the lack of a clear rule or pattern (driven by her now current meaning for randomness), the inability to state what will happen next. Implicitly, her image of stochastic process contains the ideas of replicability, but she has yet to fully explicate those ideas for herself. Finally, she was able to build a connection between random/stochastic processes and the idea of getting data that afforded her an opportunity to move beyond viewing data as fodder for producing data visualizations.

Bonnie's Second Session

During Bonnie’s second exploratory teaching session, I asked her to recap what she had done in the first session and then had her work through the activities of the Sequences Applet. Bonnie did not revert to the independence-esque meaning for randomness she gave in her clinical interview. Rather, she conveyed an image that randomness was about the lack of a pattern or rule (line 3 of Excerpt 32). What stands out the most in Bonnie’s description of the previous session is her characterization of the process categories (line 7). The distinction between the two categories of processes appears to be driven by her image of randomness. Things that are driven by some rule and therefore aren’t random come from one type of process; when there is an absence of a rule (i.e., there is randomness), then there is a lack of control over such a process.

Excerpt 32. Bonnie’s recap of the first exploratory session

1	Bonnie	Yes. So, the last time, well with the computer we had four different rooms and we were seeing basically if there's almost like a pattern within turning on the lights and we created a new light switch for each room and umm, the first room, the first question, or one of the first questions you asked me was "Where does data come from?" And then after we did all the examples of the four rooms, it kind of summed up how it's either just like super random or...
2	Neil	What's "super random"?
3	Bonnie	Umm, information, how it's, ummm, what's the word, how it's...when you go out to go get information or data. Some of it can be random, how...like we said if you went to go stand on the corner of Rural and Apache at the same time every Monday, well there's going to be random cars. Or how, ummm, when you have me turn away and look and if your shoe was still the same shoe there, it wasn't then random, it was like set by a rule.
4	Neil	Okay. (coughs) So we worked with the four rooms [Bonnie: mmm-hmm (agreement)] and each of those four rooms had a button that you mentioned before, we sort of created a process so that you could create a new button [Bonnie: Yes], right? And then I gave you a few new processes [Bonnie: mmm-hmm] such as standing on Rural [Bonnie: yes] and counting red cars. [Bonnie: mmm-hmm] You ended up making how many categorizations of processes?

5 Bonnie Ummm, I believe I made two.
6 Neil Do you remember what those two were?
7 Bonnie Ummm, one of them was categorized as there's like a set rule almost for it, how if you do one thing, it is going to affect the next, umm, factor. But then in the other category, it was just completely ummm, just random and wack out of control.

The Sequences Applet consists of fourteen processes (eight are deterministic, six stochastic) where the user only has access to the end result of up to forty trials. The intent behind this applet is to have the user try to distill which category the generating process belongs to when he/she only has access to the outcomes. Process Zero is a training process to help the user get comfortable with the applet. Bonnie quickly moved through this process, noting that Process Zero had a set rule (her name for deterministic) of adding one to the previous value, starting with one. Following this, Bonnie moved on to Process One (Excerpt 33).

Excerpt 33. Bonnie classifies the first process

1 Neil We're going to do Process One. [Bonnie: okay] (Neil clicks) [Bonnie: it's a one] We got a one. Do you have enough to decide what's going on here? [Bonnie: nnn, not yet] (Neil clicks again)
2 Bonnie (softly) okay, okay, umm, can you hit it again [Neil: mmm-hmmm (clicks)] Ooo!
3 Neil So why ooo?
4 Bonnie I was expecting a three.
5 Neil You were expecting a three? [Bonnie: yes] Why were you expecting a three?
6 Bonnie Umm, well my thought was that if it was on the process of one it should just continuously go by one but then that's how was the zero process was. So, ummm, I'm not sure why it's, there's, Oh! Can you hit it again actually?
7 Neil What was your "oh"? I'll hit again after you've your oh. [Bonnie: So, okay]
8 Bonnie Umm, I had a thought, if you do one and since there's two it would be like two two, I'll write it out [Neil: okay] I don't know if it's right, but...so you have just one, and then since there's two, two two, and then for three, there'd be three of them (writes 1, 2, 2, 3, 3, 3,) [Neil: okay] So, I'm not sure if that sounds good,

9	Neil	So I'll click once more for you. [Bonnie: (softly) alright] And again [Bonnie: (softly) yeah]
10	Bonnie	And one more time. Okay.
11	Neil	So should I click one more?
12	Bonnie	Yeah. Okay. Umm...
13	Neil	You seem very pleased with yourself.
14	Bonnie	Yes, I am. I think it should be categorized as a set rule process.
15	Neil	And why's that?
16	Bonnie	Ummm, well to me it makes sense if you one, then two well you're going to write two of them, and three, you're going to just keep increasing as the number increases [Neil: okay] to show the rule...

While she originally thought that Process One was going to mirror Process Zero (lines 2–6), she quickly hit upon the idea that each number was getting listed the same number of times as the numbers value (lines 8–16). Bonnie felt very confident with herself. I had a single four showing in the applet's window when I asked her how many fours she expected to see as in Excerpt 34. With her confidence shaken in what she believed to be the rule of Process One, she persists in attempting to figure out the rule (line 7), but she brings up the possibility that Process One is random (line 9).

Excerpt 34. Bonnie gets stumped by process one

1	Neil	So I have one four showing right now. How many fours should I have?
2	Bonnie	Four.
3	Neil	Four. So, if I click three more times, they should all be fours [Bonnie: yes] (Neil clicks)
4	Bonnie	Oh no! (sighs) Okay...can you click one more time? (Neil clicks) One more time (Neil clicks)
5	Neil	So are you still sure or not sure anymore?
6	Bonnie	Hmmmm, I'm not sure anymore. [Neil: Okay] Umm, can you hit it one more time [Neil: sure] mmmm, so what are you thinking?
7	Bonnie	I'm not sure...I'm either thinking that it could just be random or how, so for the, how the one was just grouped as one and the two is grouped as two, and three was grouped as three, but then the four went back to, went back down to just one four, and then we went to five and there's now four fives. So, then my next prediction would be that there would be five sixes and then one seven.
8	Neil	So five sixes. So, we, so the next click should start a six?

9 Bonnie Yes. (Neil clicks) Oh no. Okay...I don't know. [Neil: You don't know?] Ummm, I think it should be random then. [Neil: so, you're thinking random] Yeah.

10 Neil What if we click again? [Bonnie: we can try] Do you have a guess? [Bonnie: I still think it is going to be six] (Neil clicks)

11 Bonnie Now it is. One, two, three, four, five (counting the fives). Umm. So, okay, so you know how there's two twos, three threes, and then four, there's just a single four [Neil: mmm-hmm] but then it went, it counted out as five fives, so then if there's, there could be possibly six sixes and then it would just go back to one seven. So, it is kind of is repeating a pattern where it goes from one and then a sequence of, or like a section of two twos, and then a section of three threes, and then back down to just one number to represent four.

12 Neil Okay. So, if I click "run processes once", what should I see?

13 Bonnie A six. (Neil clicks) Oh no, okay I think it is random.

14 Neil You think this is random? [Bonnie: yeah]

15 Bonnie All of my like predictions have been turned down, (laughs), ummm, yeah I think that it's random.

After Bonnie makes a several more guesses and tests out a few more potential rules (lines 11–15), Bonnie has consigned Process One to her category of random processes. Given that her image of random consists of the lack of a rule/pattern, Bonnie’s behavior can be described as pattern hunting. As she views more and more outcomes, she looks for any type of term-based pattern that would give her a set rule. When she cannot find such a rule, she concludes that the process must be random. In Excerpt 35 I proved Bonnie with the first forty terms of Process One and prompt her if she notices any rules or patterns:

Excerpt 35. Bonnie examines the full sequence for process one

1 Neil Should we go ahead and fill the board? [Bonnie: yeah] so what do you notice?

2 Bonnie Umm, there's a lot of elevens.

3 Neil How many elevens?

4 Bonnie (counts) Eleven. [Neil: eleven]

5 Neil How many sevens are there? [Bonnie: seven] How many eights? [Bonnie: one] Nines? [Bonnie: one] Tens? [Bonnie: two, errr, one]

		Twelves? [Bonnie: one] And for thirteens, we have [Bonnie: five] five, but we're out of terms [Bonnie: mmm-hmm]. Right, so there could be more [Bonnie: yes]. Do you notice any rules or patterns?
6	Bonnie	Ummm, let me look at it really fast. So you have [Neil: go ahead and talk out loud for me] Okay, so I'm just seeing if, again how it just starts with one and then goes two of them and then three of them and then it, for four it is just represented by the number four and then five is represented by five fives, but then six is just represented as one six and I was seeing if there's a correlation between like how many number in between these single numbers are almost like spread out, if that makes sense. And right now, I'm not seeing any rule, because between the one and the four, they're like the single numbers, there's two numbers like spread out or expanded. But the five and the seven there's only one number. And then between the seven and the eleven there's three numbers. Hmm
7	Neil	So is there anything special about the numbers that aren't repeated?
8	Bonnie	That aren't repeated...hmmmm, I don't see anything.
9	Neil	You don't see anything? [Bonnie: nnnnhh, not yet, hmmm]
10	Bonnie	(softly) I don't...I don't see a pattern. [Neil: okay] So I think that it's random.

Even given my prompt, Bonnie engages in pattern hunting behavior in line 6. Her focus is one looking for some pattern to the spacing values: how many non-repeated numbers appear between repeated numbers and how many repeated numbers (not terms) occur between non-repeated numbers. I attempted to prompt Bonnie to look at the numbers that weren't repeated and see if there was anything special about them. She did not see anything and concluded that Process One was random.

At the end of the session, I asked Bonnie to return to Process One. I explicitly told Bonnie that the numbers that weren't repeated all had something in common and all of the numbers that were repeated had something in common. After writing down the two lists of numbers, Bonnie recognized that the list of repeated numbers contained the primes. She was then able to use this information to change her classification of Process One from random to set rule.

Excerpt 36. Bonnie classifies process two

1	Bonnie	Alright, let's go for another. (Neil clicks) One more. Is that a G? [Neil: that is a G] Okay. Another one (laughs) G. One more. Hmmm
2	Neil	What's the hmmm?
3	Bonnie	Okay. So, you start with A, you skip B and you go straight to C, but then you skip a whole bunch of others (both laugh) for G, and there's two Gs in the process too. And now there's two Cs, so maybe there could be an A next?
4	Neil	Should we try? [Bonnie: yeah] (Neil clicks, gets an A) [Bonnie: okay]
5	Bonnie	So, what I'm thinking right now is that we started with A, then we skipped B, which I don't know why, but or may because it is Process Two, so you skip...but you'd only be skipping one though
6	Neil	So the process names don't have [Bonnie: okay] anything other than they are just [Bonnie: oh, okay, okay]. Yep.
7	Bonnie	So, okay, hit it again. (Neil clicks) A. Hit it one more time. (Neil clicks) And one more time. (Neil clicks, gets T) Hmm.
8	Neil	So what are you thinking?
9	Bonnie	Ummm, I thought it was going to be a pattern of A C G G C A, and then since there is another A, that it would be the same pattern again. Because another C came after but instead of G it's T. So, I'm not sure now. But I think that there will be another T after this.
10	Neil	Shall we try? [Bonnie: yes] (Neil clicks, gets A)
11	Bonnie	Oh no, okay. Umm, so the reason why I thought there would be another T is because in the first few of them, it went from A C G G, so then I thought it would be going A C T T, but instead it went A C T A. One more time. (Neil clicks) Back to C. Okay. Hmmm, I don't see a pattern.
12	Bonnie	Can we go one more time? (Neil clicks) Okay. Yeah, I don't know if it, I think that this is just a random proc--errr, just a random.
13	Neil	And why's that?
14	Bonnie	I don't really see any pattern or fixed rule, that the letters are following, it's kind of just being thrown in there, one after another, just randomly.
15	Neil	Should we fill the grid? [Bonnie: yes]
16	Bonnie	Hmmm, so I don't really see any same pattern as the first line.
17	Neil	So none of these other three lines don't mimic the very first line? Right?
18	Bonnie	No. Or, it doesn't really look like they mimic each other at all. Just kind of just all thrown out there. So, yeah, I think that that one's random. [Neil: okay]

Excerpt 36 shows Bonnie's work with Process Two. She was initially surprised by Process Two giving her letters, but she quickly moved past that and engaged in pattern hunting. In line 3, Bonnie picks up on the potential for a palindrome in the outcomes (i.e., ACGGCA). When I clicked the applet for the next term, the applet did give "A" to complete the palindrome. On subsequent clicks, Bonnie continued to search for her palindrome pattern (lines 7–10). However, her rule breaks in line 11 causing her to abandon her pattern hunt and classify process two as random. Even after filling the grid (lines 15–18), Bonnie still searched for her palindrome pattern and then for any sign of a pattern. Bonnie's pattern hunting scheme in Processes Zero through Two is her only approach to deciding what type of process is at play. However, her scheme is limited by what she can recognize as making up a pattern. Her explorations with Process Five (Excerpt 37) highlight the types of patterns she looks for: additive and positional. In lines 5–9, Bonnie looks for an additive pattern by looking at the differences in successive terms. When she does not find a pattern to these differences, she concludes that the process must be random (line 17).

Excerpt 37. Bonnie classifies process five

1	Bonnie	That's a random number (referring to the –0.68 showing) (laughs)
2	Neil	Why do you think that's a random number?
3	Bonnie	Umm, I don't, who starts counting at 68, especially negative point 68? (laughs) It's an even number. Alright [Neil: that is an even number] (laughs)
4	Neil	Shall we go for another one? [Bonnie: yes] (Neil clicks, gets -0.36)
5	Bonnie	Okay, I'm just seeing if there is any correlation here, or relationship, ummm...so far I don't see any except that it's even. [Neil: Okay (clicks again and gets 0.44)] And now it's positive but it's still even. Ummm, [Neil: ready?] yeah. (Neil clicks and gets 0.79) Umm, I'm seeing, or I'm thinking if 44 plus, what would it be, 35 if that's the same amount it takes for 68 minus 35...but it's not. [Neil: okay] (Neil clicks again and gets 0.84) Hmmm,

6	Neil	What's your hmmm?
7	Bonnie	So, we went from point 44 to point 79, which was a pretty decent amount, but then from point 79 to point 84 it's about 5, different, err an addition of 5. [Neil: okay] So, I'm not really sure if there's a pattern so far.
8	Neil	Shall we go again?
9	Bonnie	Yes. (Neil clicks and gets 0.61) Now it went back down. Yeah, so far I'm not seeing any rule.
10	Neil	Okay. Shall we go again? [Bonnie: mmm-hmm] (Neil clicks and gets –0.66)
11	Bonnie	Now it's negative. Okay. Can we go one more time? (Neil clicks; –0.4) Really negative. Okay
12	Neil	What do you mean by really negative?
13	Bonnie	I didn't mean really negative; (both laugh) I just read that wrong. Ummm, I don't see anything now, right. [Neil: nothings jumping out at you?] Hmm-mmm (negative). Except that we just went, we started at negative numbers, went to positive, and now we're back down to negative numbers.
14	Neil	Shall we go again?
15	Bonnie	Yes. (Neil clicks) Yeah, I don't know. [Neil: You don't know what?] If there is a rule. It seems, it's their just random numbers being thrown out. [Neil: okay (clicks again)] Yeah, I think it's random. (laughs)
16	Neil	So you think this is random. Why do you think this is random?
17	Bonnie	Umm, how it went from, from negative point 66 to negative point 4 and then back down to negative point 85 and then really jumped to point, negative point zero seven, there's really just not a rule for that. It's just going like up and down [Neil: okay; shall I fill the grid?] Yes.
18	Neil	So what are you noticing?
19	Bonnie	Ummm, the first thing that caught my eye, I don't know if it is just a coincidence, but here's it's negative point 36 (points to second box of first row) and here's its negative 1 point 36 (points to second box of fourth row), uhh. [Neil: does that work anywhere else?] Well, here it's one point twelve (points to third box in fourth row) and uhh, that doesn't make sense, I was going to say that twelve times twelve is 144, but that's just zero point 44 (points to third box of first row). Umm, I don't see any patterns. Hmm
20	Neil	So the process that generated this sequence [Bonnie: mmm-hmm] how would you categorize this process? [Bonnie: random] Random? [Bonnie: yes] And why's that?
21	Bonnie	Ummm, I just don't see any set rule of going up or down a certain amount of times or multiplying or dividing a certain amount of times.

After filling the grid, Bonnie switches to her other pattern type: positional. Her last positional pattern would be the palindrome she saw in Process Two. Here, she's looking across the rows to find some type of pattern (line 19). She attempts to make a connection between terms in the same position but different rows starting with a –0.36 and a –1.36. However, Bonnie gives up this approach and concludes that Process Five must be random since she cannot find a set rule (line 21).

Process Seven provided Bonnie with an interesting experience. Process Six had clued her into thinking about standard, six-sided dice, which she brought up in early on when working with Process Seven as in line 1 of Excerpt 38. For her, the invocation of a die immediately indicated that the process was random. In line 5, Bonnie used her positional pattern hunting scheme as she saw a sub-sequence (3-3-2-6) appearing again. However, when that particular pattern failed, I believe she switched to looking for an additive pattern (line 9) before switching back to searching for a positional pattern (lines 11–15).

Excerpt 38. Bonnie classifies process seven

1	Bonnie	A three again. (Neil clicks) Another 3. (Neil clicks) A two. [Neil: Again?] Yeah. (Neil clicks; 6) One more time. (Neil clicks, 1). One more time. (Neil clicks, 2) So now that you brought up the die, it seems like this could be random as well.
2	Neil	So what will we get next?
3	Bonnie	What was that?
4	Neil	What will we get next?
5	Bonnie	Oh, ummm, maybe a six. [Neil: you don't sound so sure] Umm-mmm. (Neil clicks) Three. Okay. One more time. (Neil clicks, 3) One more time. (Neil clicks, 2) And then there would be a six. (Neil clicks) Oh, one. Okay. Umm, so then after I thought I was starting to see some type of rule after, cause we started with a three and then we had a three and two, so then once we had another three and another three after that and then a two, I thought "oh it's going to be the same" rule, pattern, but then we got a one instead

of six. [Neil: okay] So, kind of threw it off almost. One more time, can you hit it? (Neil clicks, 1; 5) Yeah, I think that it is random.

6 Neil So you think this is random? [Bonnie: yeah] Shall we go again? [Bonnie: Yes] (Neil clicks, 2)

7 Bonnie One more time. (Neil clicks, 6)

8 Neil So what are you thinking about?

9 Bonnie Ummm, well, I was just thinking that the five and the three, err, two and the three is five, but I don't really see relationship with those numbers. Hmmm, yeah, I don't see anything. [Neil: okay. Shall we go again?] Yes. (Neil clicks, 1) The only thing is that in the first row we had a six and then a one; and we have another six and then a one. Oh, and a two in front of that six.

10 Neil So what would come next?

11 Bonnie A two. (Neil clicks, 2). One more time? (Neil clicks) A six. I was expecting a three.

12 Neil Why were you expecting a three?

13 Bonnie So in the first few, after we did two six one two, there was a three. But then now there's a six.

14 Neil So what do you think will come next?

15 Bonnie A six. (Neil clicks, 3) Oh, three. Okay. [Neil: so, what do you think will come next?] A three. (Neil clicks, 2) Oh no, a two. (laughs) Yeah, I don't see any rule or pattern here.

16 Neil You see no rule or pattern. [Bonnie: mmm-nnn]

Bonnie had yet to pick up and comment on the reoccurring twos, so I filled the entire applet's gird (giving all forty terms). Even with the filled grid, Bonnie still did not notice anything about the twos (line 2 of Excerpt 39). I found this odd given her prior positional pattern hunting. I sorted the outcomes and Bonnie then noticed all of the twos (lines 4–6). With my heavy prompting, Bonnie was able to describe that a two would appear every third term (line 8).

Excerpt 39. Bonnie notices a lot of twos

1 Neil What if I go ahead and fill the board? [Bonnie: alright]

2 Bonnie Hmmm. So far, I'm just seeing...this is still the same numbers one through six, umm, I'm seeing...hmmm...I just don't see any number, anything that catches my eye that could be really a rule or pattern. I think that this one is random.

3 Neil I'm going to go ahead and sort this [Bonnie: okay] What do you notice?

4	Bonnie	There's a lot of twos, like a lot.
5	Neil	A lot. [Bonnie: yeah] Like more than you would expect? [Bonnie: yeah] Why are there more than you would expect?
6	Bonnie	Umm, random that has so many twos. Cause when I first counted one, there's only four, so I was expecting well maybe there will be either four twos or five twos, but instead there's a whole group of them [Neil: 13 of them] Yeah. Thirteen doesn't seem like a relevant number in this situation.
7	Neil	Well, let's look back at the original ordering [Bonnie: okay] maybe there's something going on with the twos.
8	Bonnie	So after every two numbers, there's a two. So, [Neil: Always?] Yes. You start with these two (points to first two boxes), there's a two; these two, there's a two; these two, two; these two, two. It seems to be there's always a two following two numbers. Yeah.
9	Neil	So we ended on term 40 [Bonnie: mm-hmm]. If we go to 41, what would be the 42 term?
10	Bonnie	A two. [Neil: A two] mmm-hmmm [Neil: guaranteed?] Yeah.
11	Neil	So, is there a rule or is this a random?
12	Bonnie	Hmmm, oh goodness, okay, I'm going to, so, I think that over all the numbers are random, but how's there a two after every two numbers doesn't seem random to me. So, I don't know if that overrules the other randomness. I think that it would be under a set rule.
13	Neil	So you're thinking that there is a set rule? [Bonnie: yeah] So you would [Bonnie: just in a different way] So different from...
14	Bonnie	So, usually if there is a set rule, you know you would maybe have three threes and then like four twos and five sixes, like some, kind of like going up by the same amount, or going down by the amount, but then here we always know that after every two numbers, the number two will be there. So, it's like a reoccurring, umm, factor.
15	Neil	So rather than the set rule telling us every value [Bonnie: yeah] this will only tell us [Bonnie: every] every...third [Bonnie: third value] value.

Bonnie struggled with classifying Process Seven (lines 12–15). Given that she had a set rule for every third term but not the other terms, Bonnie ended up classifying Process Seven as both Set Rule and Random. I designed Process Seven to violate von Mises's Principle of the Impossibility of a Gambling System and thereby not qualifying the process as stochastic. Strictly speaking, Process Seven cannot be a stochastic process since the process lacks the appearance of randomness. However, Process Seven does not fit cleanly within the notion of a deterministic process either.

The last of processes from the Sequence Applet I'll discuss in relation to Bonnie is Process Ten. Process Ten involves the Fibonacci Sequence, rearranged so that the second half of each set of eight terms appears first, then the first half, then on to the next set of eight terms. Bonnie immediately engaged her additive pattern hunting scheme by looking at successive differences (line 1 of Excerpt 40).

Excerpt 40. Bonnie classifies process ten

1	Bonnie	(Neil clicks, 5) Okay (Neil clicks, 8) Okay. (Neil clicks, 13) So, from 5 to 8 we went up by three, and from 8 to 13 we went up by 5, right? [Neil: mm-hmm (agreement)] Yeah. Can you hit it one more time? (Neil clicks, 21). So, we went from 3 to 5? [Neil: we went from 3 to 5?] So, from adding, so you 5 plus 3 is 8, 8 plus 5 is 13, so we from 3 to 5. [Neil: okay] So, then, 13 to 21 would be eight, so then you went from 3 to 5...so from 3 to 5 you added two to that, and then from 5...I just lost my train of thought. Okay so 5 plus 3 is 8, and 8 plus 5 is 13, oh, okay, 3 plus 2 would be 5, so you would go up by 2, then 5 plus 3 would be 8, and that's, how, what, the amount you went up by next. So then, like you're going to continuously one to the number you're adding by. So, the next you would add by...12.
2	Neil	So what would we get next?
3	Bonnie	(softly) (unintelligible) 5 plus...3 is 2...umm, I think you should 32.
4	Neil	Are you ready? [Bonnie: yeah] (Neil clicks, 1).
5	Bonnie	Oh no! (laughs) okay. Ahhh, can you hit one more time? (Neil clicks, 1) One more time. (Neil clicks, 2) Ahh, I don't, I, ehh, that's random. (laughs)
6	Neil	This is random? Why do you think random?
7	Bonnie	Umm, I don't see any relationship going from 21 back down to 1 and then 1 again and then 2.
8	Neil	Shall we go again?
9	Bonnie	Yeah. (Neil clicks) And then 3. One more time? (Neil clicks, 233) (sighs) Hmmm, I don't know. I have no idea (laughs) [Neil: No idea what's going on?] mmm-hmmm, I don't think that there's any rule. (Neil clicks, 377) Hmmm, I mean so far the only numbers used are, well 2 and 3 are used a lot, there's a few ones in there. [Neil: shall we go again?] Yeah. (Neil clicks, 610) Hmmm...umm,

I'm not, I don't see anything so far. (Neil clicks, 987) Well, kind of. [Neil: What's kind of?] If you think about, if you're starting at, umm, 300 plus 300, essentially get 600. And then 600 plus another 300 would essentially be 900. [Neil: mmm-hmm] But, I don't know where...

10	Neil	Shall we click again?
11	Bonnie	Yes. (Neil clicks, 34). Awww, we got...I think it's random.
12	Neil	You think this is random? [Bonnie: yeah (laughs)] Shall we fill the board? [Bonnie: yeah]
13	Bonnie	Oh yeah, I think that's random [Neil: you think this is random?] Yeah. [Neil: do you notice anything going on with the numbers?] Umm, they increase, then they go back down, then increase again, ummm,...(long pause)...umm, I don't know if this is a factor but they're kind of like in groups of four. So, this (points to the first four) is increasing when there's four numbers, and this one (second set of four) like you go back down but you're increasing like another four numbers, and then these four numbers (points to third set of four) are like increasing. And then these (points to fourth set of four) ones you go back down but you're still increasing.
14	Neil	So they might be in groups of four [Bonnie: yeah].

Setting addition errors aside, Bonnie was able to detect an additive pattern to the differences of successive terms (lines 1, 3, and 9) and see was able to pick out the positional pattern of sets of four terms (lines 13 and 14) but she could not bring these two patterns together. This suggests that for Bonnie, patterns are either one type or the other, without any blending. Due to this issue, Bonnie concludes that Process Ten must be random. In an attempt to get her to focus in on the Fibonacci sequence, I proposed using the applet's sorting feature (Excerpt 41). In the sorted form, Bonnie lost track of the prior additive pattern she had (line 2). What is interesting here is that when I prompted her in line 3 for the classic Fibonacci sequence, she was able to see that pattern but made no connection back to her original additive pattern of successive differences (lines 4–6). Her original additive pattern (Excerpt 40, line 1) focused on how the successive differences where increasing in a fixed way but she made no connection to the prior terms. With my prompt (line 3 of Excerpt 41) she was no longer thinking about successive differences but rather the addition of the prior two terms. Now armed with an additive pattern to serve as a rule, Bonnie classified the sorted version of Process Ten as being a set rule (line 8). I then brought back the positional pattern (line 9).

Excerpt 41. Bonnie struggles to coordinate two types of patterns

1	Neil	Well, what if we sorted this list? So put everything in numeric order? (Neil sorts the list) Now do you notice anything?
2	Bonnie	Ummm...(long pause)...umm, I don't notice anything too crazy. [Neil: what do you mean by too crazy?] I don't notice any rules that stand out to me, except that there's two ones. [Neil: what comes after those two ones?] One two. [Neil: what comes after the two?] A three. And then a five. And then an eight. (long pause). Ahhh, I don't know [Neil: you don't know?] Hmm-mmm (in the negative)
3	Neil	What if I said, what's the relationship between the third term and the prior two?
4	Bonnie	Two plus one is three. [Neil: third term; prior two] Oh, one plus one is two. [Neil: what about the fourth term?] Two plus one is three. Ohhh! So, you, you start by taking, by looking at the third term and the two before that term you add those together to get that term. [Neil: does that pattern hold?] I think so. Yeah.
5	Neil	So now let's look at the original ordering. [Bonnie: okay]
6	Bonnie	So if you look at 13, 5 plus 8 is 13. 8 plus 13 is 21. Umm, but then once you get to the... (long pause) so then, I don't think it holds. [Neil: you don't think holds?] Mmm-mmm (negative)
7	Neil	So we had this sequence sorted [Bonnie: yeah] like we do now. What type of process would you say generated this?
8	Bonnie	Having a set rule.
9	Neil	Having a set rule. But if we look at the original order, now what would you say?
10	Bonnie	Well, it works for like the first few numbers, so, when we say by the third number, this is where I'm confusing myself. So, you know you look at 13, 5 plus 8 is 13, so then would you look at 21 and say 8 plus 13 is 21? [Neil: yes]. Okay, so then if you look at one, well 21 and 13 doesn't equal one. [Neil: okay] But then if you look at 2, 1 plus 1 is 2. If you look at 3, 2 plus 1 is 3. But 2 and 3 do not equal 233 [Neil: Yeah, but what do 2 plus 3 equal?] Five. [Neil: is there a 5?] Umm, in the next number? [Neil: not, just is there a 5?] There's a 55. [Neil: there's a 55, isn't there a 5 back here?] Oh yeah.

11	Neil	What's 13 and 21 added together?
12	Bonnie	Uhh, what is, 34. [Neil: is there a 34?] Yeah. And then there's 55, then 34, 89...so there is a rule, but it's not...it's like, [Neil: it's not the same] same order. It's a rule that's kind of scatter throughout the board but doesn't follow specifically a pattern.
13	Neil	Is the rule scattered?
14	Bonnie	Hmmm
15	Neil	So earlier you talked about how there's this group of four and there's this second group of four, and there's a third group of four [Bonnie: mmm-hmm (agreement)], and then there's a fourth group of four [Bonnie: yeah].
16	Bonnie	So, then if you look within those four-groups, then the pattern would work I think.
17	Neil	And if you notice if you take the second group of four and then go to the first group of four [Bonnie: mmm-hmmm] and then you the third, the fourth group of four
18	Neil	[Bonnie: they still] and then go to the third group of four [Bonnie: yeah] and then you would have to the...
19	Bonnie	Sixth group and then back to the fifth. Okay. So, then there is a pattern. [Neil: there is a pattern, just not quite the same type of patterns we had before] Yes.
20	Neil	So would this be a random process?
21	Bonnie	No [Neil: why?] Umm, again, if we can go back and determine what we know with a specific pattern or rule, it's not going to be random.

Bonnie struggles to coordinate both the additive pattern and the positional pattern to Process Ten (lines 10–14). Bonnie sees the additive pattern working for the first three terms and breaking down with the fourth since 13 plus 21 does not equal 1 (the fifth term). When I specifically point out the groups of four, Bonnie notices that the additive pattern holds. I then gave Bonnie the rest of the positional pattern (lines 17 and 18), which she was then able to carry forward to additional sets of four (line 19). This ultimately led Bonnie to declare Process Ten as not being a random process but as having a set rule. However, I do not believe that Bonnie could have arrived at this decision without my explicit intervention of coordinating both the additive and positional patterns for her.

Table 11. Bonnie's Classifications of the Processes in the Sequence's Applet

Process	Bonnie's Classification	Actual Classification
Zero	Set Rule	Deterministic (n)
One	Random, then Set Rule	Deterministic (primes are listed their numeric value number of times; others once)
Two	Random	Stochastic ($\mathcal{DU}(4)$; {A, G, C, T})
Three	Set Rule	Deterministic ($2n + 1$)
Four	Set Rule	Deterministic ($n^2/100$)
Five	Random	Stochastic ($\mathcal{N}(0,1)$, rounded two decimals)
Six	Random	Stochastic ($\mathcal{DU}(6)$; {1, 2, 3, 4, 5, 6})
Seven	Random & Set Rule	Non-stochastic (2 if $n \equiv 0 \bmod 3$, otherwise $\mathcal{DU}(5)$, {1, 3, 4, 5, 6,})
Eight	Set Rule	Deterministic ($n - \sin(45)$, rounded two decimals)
Nine	Set Rule	Deterministic (–6)
Ten	Random, then Set Rule	Deterministic (Fibonacci's Sequence, rearranged in blocks of four)
Eleven	Random	Stochastic ($\mathcal{E}xp(2)$, rounded two decimals)
Twelve	Set Rule	Deterministic ($-1/n$), rounded five decimals)
Thirteen	Random	Stochastic $X(n) = X(n-1) + err(n)$ where $err(n)$~$\mathcal{N}(0,25)$)

Note: The names "set rule" and "random" are Bonnie's names for the categories she created in the prior session using deterministic and stochastic processes, respectively.

Table 11 shows Bonnie's classification of the fourteen processes in the Sequence Applet as well as the actual classification. For the most part, Bonnie's pattern hunting scheme allowed her to make normatively correct classifications. However, there are some notable exceptions: Processes One and Ten. During her initial time with Process One, Bonnie ended up classifying the deterministic process as being random. I believe that given her pattern hunting scheme being focused on additive or positional patterns, she was not disposed to look for other kinds of relationships. She was only able to make the normative classification after I prompted her to notice that the repeated numbers were prime. Process Ten held both additive and positional patterns; however, Bonnie struggled to coordinate the two. Bonnie initially saw an additive pattern in the successive differences but could not carry that pattern into successive blocks of four terms. However, she did detect that there were blocks of terms. Only after I supplied the coordination of the two patterns was Bonnie able to make the normative classification of Process Ten.

Process Seven represents an interesting case. In the prior sessions, I did not provide technical definitions for either deterministic or stochastic processes. Nor did I insist upon them in the present session, relying instead on her own classification system. However, the nature of Bonnie's struggles with Process Seven speak to the issues she had with Process Ten. Specifically, looking at the positional pattern of the twos. Until I had sorted the outcomes, Bonnie did not notice that every third term was a two. This suggests that her positional pattern hunting is limited in scope to looking for repeated sub-strings (more than one character) or looking for chunks such as palindromes. In Process Ten, Bonnie noticed chunks of four terms, but could not see what to do with these chunks as they did not fit within repeated sub-strings nor were they palindromes.

Bonnie's pattern hunting scheme drove her work, in particular her additive pattern hunting. She used the such a strategy with every process except Processes One, Two, and Nine. For the most part, she looked at the differences of successive terms and built a rule centered on those differences; if I supplied her an alternative rule such as squaring or doubling, she would accept the proposal. While I do not know her full mathematical background, a common aspect of many American students' experiences with school mathematics is that of looking for patterns in a sequence of numbers. Often students are taught to look at the differences between successive terms and even between successive differences. This is what Bonnie naturally did each time she hunted for an additive pattern. Processes Three and Four both have multiplicative aspects (doubling and squaring, respectively), but she still worked additively. Unfortunately, the Sequences Applet does not have a process that is multiplicative in nature and where additive strategies won't work, for example, exp(x). I suspect that Bonnie would struggle to classify such a process and would potentially call such a process random.

Bonnie's image of randomness as being without a pattern or rule served as the ultimate driver for her classification process and drove her to pattern hunt. Her past experiences of looking for patterns limited her in what kinds of patterns she was pre-disposed to look for.

Chapter 8: Danielle's Meaning for Stochastic Process

After her initial clinical interview, Danielle came in for two additional sessions of a teaching experiment. In the first session, she worked through the Light Switch (Version A) and the More Processes Activities (see Chapter 5: Methodology). At the conclusion of the More Processes activity, Danielle worked through the intervention shown in Figure 20. At the start of the second session of the teaching experiment, I provided Danielle with a similar sheet as what she filled out at the conclusion of the first teaching experiment session. My intent was to provide Danielle with a tool that could help her explore the processes behind the Sequence Applet and solidify her meaning for stochastic process.

Teaching Experiment Session One

As with Bonnie, I asked Danielle if she had ever heard of a stochastic process before and where data comes from at the start of the session (Excerpt 42). While she claims to have heard the phrase "stochastic process" before, she can't recall what that refers to or where (Lines 2 and 4). In contrast to Bonnie (data appears, and you make visualizations), Danielle's initial answer to where data come from contains an explicit reference to going out to collect data and then manipulation coming later (Lines 6 and 8).

Excerpt 42. Danielle's initial view of stochastic process and the origin of data

1	Neil	Have you ever heard of the phrase "stochastic process" before?
2	Danielle	Yes, I've heard of it. But to tell you what it is, (I) wouldn't even begin to remember
3	Neil	Okay. Do you remember where you heard the phrase?
4	Danielle	Ummm, no. Since we're here, I'm assuming that it's a statistical thing [Neil: laughs] but that's just me.
5	Neil	That's a good assumption, right. [Danielle: laughs] So another question I have for before we get started, or really started with the applet [Danielle: umm, kay]. Where do data come from?

6	Danielle	Data? [Neil: mmm-hmm] Umm, numbers of people on the ground, crunch for whatever reason.
7	Neil	But where do they get those numbers to crunch?
8	Danielle	Well, in my case, we go out and we count plants. And then we tell our supervisor, this is how many plants we counted [Neil: laughs, okay] So...then he has to deal with the hundreds of plants we count by the end of the season.

Danielle worked through Room One (deterministic, white and black) rather quickly. When asked to create a new switch to replace the “broken” one for Room One, she creates the following diagram (Figure 31) and explains that her switch will toggle back and forth between on and off—depending on where the room is currently with each press of the switch (Excerpt 43).

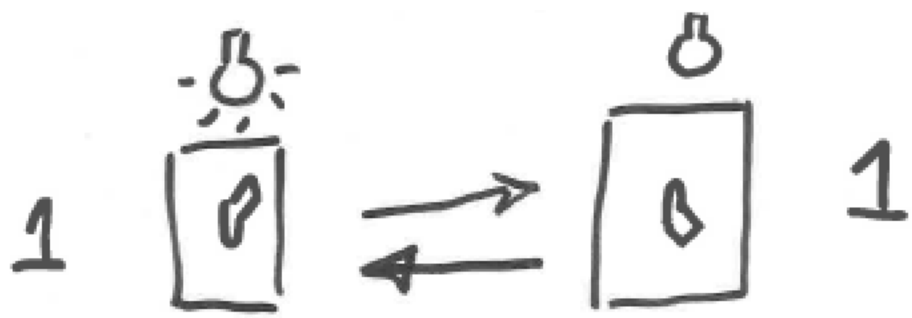


Figure 31. Danielle’s new switch for room one

Excerpt 43. Danielle’s rules for room one's new switch

1	Danielle	So, I would need to have a section for turning it on and then turning it off [Neil: okay, if that's what you so choose] (laughs) Alright.
2	Danielle	So, I'll just go with a classic switch [draws a light switch in the on position] and then [draws an arrow; writes "1" to the left of her first switch;] oh I messed that there [draws a second switch in the off position; adds on and off lightbulb above]
3	Neil	Okay. [Danielle: alright] So describe for me what you've drawn.
4	Danielle	Umm, so, if you hit the switch once, if it is on, it will turn off.
5	Neil	Okay. So, if we hit your switch right [hits switch]
6	Danielle	Okay.
7	Neil	So if I hit the switch again...
8	Danielle	[adds another arrow in the opposite direction and a 1 to the right of the second switch] It will turn back on again.
9	Neil	So, can we only hit the switch once?

10 Danielle Umm, no. [Neil: okay] You can hit it more than once... [gestures back and forth between her switches] (softly) back and forth
11 Neil It will just keep going back and forth? [Danielle: yeah]

In the Four Rooms activity, I ask each student who to get back to a prior state of the room after I come through and press the switch. Danielle leverages how she answered this question (correctly noting that she would need to click the button nine times and that would work for any case) to help her design the new switch for Room Two (deterministic, grey scale). In Line 2 of Excerpt 44, Danielle creates her switch (see Figure 32) based upon knowing that five clicks will get her from a black state to a white state. While she initially confounds the number of clicks with the number of states, in checking that her button works (Lines 5 and 6), she is able to quickly rectify the issue.

Excerpt 44. Danielle's description of her new button for room two

1 Neil Now, the switch breaks. (both laugh) [Danielle: alright] So, I need you to make a new switch, new set of rules, process to replace this switch [points to the screen] that no longer works. [Danielle: alright] So what would you do?
2 Danielle Let's draw it in the correct order this time, we'll start off with...completely off [draws a light switch in the down/off position with a dark light bulb above] and then [draws a second light switch in the up/on position with a lit bulb above] completely on. [Draws a wedge shape with the narrow end towards the dark bulb and the wide end towards the lit bulb, with subdivisions] [Below the switches and wedge, she draws a series of "moons" to indicate the amount light through successive stages] Close enough.
3 Neil So what are you drawing?
4 Danielle Alright, so, completely off, completely black [taps the off switch] and then [adds numbers across the top] those are off, but okay, so completely black and for every single click you'll go slightly lighter and on the fifth click, it should be completely white. And then back again.
5 Neil So, suppose that we start at completely off. [Danielle: okay] And you press your switch, or whatever, so go ahead and press your switch, your switch [points to her drawing] [Danielle: oh, that switch, okay] Yep, remember that one [points to switch on the computer screen] is broken.

6	Danielle	Okay. So, (both laugh) alright so, press the switch once it should be three-quarters of the way black, [Neil: okay] And then [Neil: if you press your switch again] Press it again, half, half of it is dark. And then, again you get three-quarters of it is light, so a quarter is dark. And then the fifth time is, crap...Alright, hold on a sec, I missed one in here. [Adds another "moon" below her switch] five times, not four times [adds a new number 5 and changes the original 5 to a 6 above her switch] Okay, so there are six stages but five clicks. There we go. Alright
7	Neil	Okay. So, we're at the stage so...if you...
8	Danielle	Yeah so, if we hit one more time, then we'll it will be white. Okay
9	Neil	And if we hit one more time?
10	Danielle	Then we'll go back down this direction (getting darker) [Neil: okay] [Danielle draws left and right arrows below her "moons"] [Neil: okay]

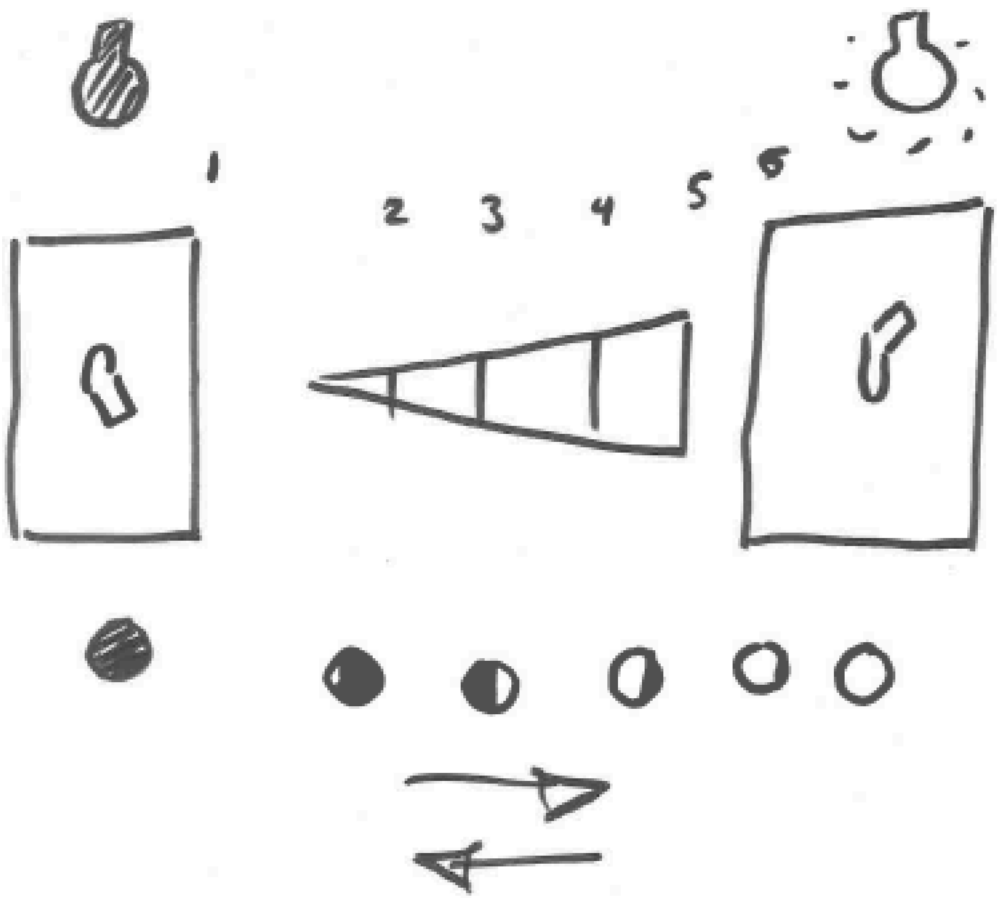


Figure 32. Danielle's new switch for room two

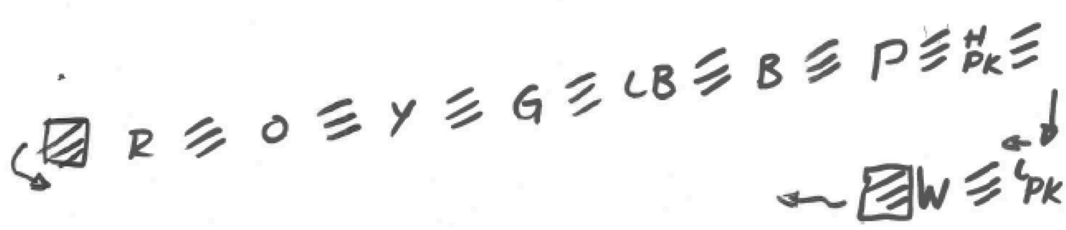


Figure 33. Danielle's sequence for room three

As in Room Two, Danielle used the question of how to get back to the previous state the room was in to build her new switch. In the case of Room Three, she wrote out the colors, using three hash marks to represent the black (light off) stages between each of the colors (see Figure 33). The boxed hash marks denote the black state that is the beginning/end of the cycle. Danielle initially struggled with the question of making a new button for Room Three; she even proposed that the new button should do a single color (i.e., behave like the button for Room One) (Line 2 of Excerpt 45). The other issue at play here was her struggle to try to come up with something like computer code to make a new button for Room Three. She did not seem to connect back to her previous two experiences making buttons. However, once she got past that block (Line 15), she was much more confident in that her new button would work.

Excerpt 45. Danielle's new button for room three

1	Neil	Switch broke for Room Three. [Danielle: okay] I would like for you to make me a new switch. [Danielle: okay] So what would you do, [Danielle points to her paper] yep. if you need more paper, [Danielle: okay] feel free to use more paper.
2	Danielle	So, in making your new switch [Neil: yep] I would pick a single color and stick with it.
3	Neil	Nope, it has to do the exact same thing that Room Three does. [Danielle: okay] (both laugh)
4	Danielle	Alright, umm, what would you have me do? So, am I supposed to get back to color it was before?

5	Neil	So, umm, there was this switch [Danielle: umm-hmm] that moved you through this cycle, right? [Danielle: umm-hmm] The switch broke, you're going to create me a new switch that would move me through this cycle. What would you do?
6	Danielle	I would make sure that the sequence is still in order [Neil: okay] Just an on/off switch.
7	Neil	So where would be the start?
8	Danielle	Probably with red. [Neil: are we starting in red] I have no idea where we're starting.
9	Neil	So, this is the room as was when you first encountered the room. [Danielle: okay] So we're starting in...[Danielle: Red.] But the room light was? [Danielle: black] So where did we start?
10	Danielle	Ohhhh, okay, we're staring back at the beginning, okay (Neil chuckles)
11	Neil	So if you start with black, then what should your button, your switch do? [Danielle: Turn it to red.] Okay. So how would you write that? In terms of, like a command, in this but-, this switch?
12	Danielle	Hmmmm, I don't know. Probably just substitute the little squiggles that I have for a "B" (for black) [Neil: okay] And then...just black, red, black orange, black, yellow...
13	Neil	So every time someone presses the switch, the switch knows to...
14	Danielle	To change to the next one in the sequence.
15	Neil	Okay. That's a command, right? [Danielle: yeah, right, computers] That's something [Danielle: (laughs) computers] that's okay. English sentences work well [Danielle: Yes, oh my gosh] (Neil laughs) Okay.
16	Danielle	Alright, so yeah, so it knows, the switch knows to go the next in the...
17	Neil	In the...? [Danielle: in the order] Okay.

From Danielle's clinical interview (Chapter 6: Students' Meanings for Randomness), Danielle viewed randomness in two ways: randomness as relating to where you know don't know why the event happened and randomness as indicating that she cannot discern a pattern. As she worked with Room Four, she brought up the notion of randomness early in her exploration (Excerpt 46):

Excerpt 46. Danielle brings up randomness

1	Danielle	Is it randomized?
2	Neil	What do you mean by randomized?

3	Danielle	Is there a pattern that I cannot discern without clicking it like 400 times?

In Line 3, Danielle reveals that her image of randomness being about not seeing a pattern to bear. The search for a pattern drives her in the question for getting back to a room with dark blue light after the room was in pink light. To tackle this question, Danielle begins to record the colors of the room as she clicks Room Four's button. Her goal for writing down the colors is to find a pattern that she could then use (Line 2 of Excerpt 47).

Excerpt 47. Danielle's quest for a pattern in room four

1	Neil	So what are you hoping to accomplish by writing out all of these colors?
2	Danielle	I'm hoping that by the time that I get to the color that I actually want, a pattern will start to form. [Neil: okay] Hoping; that's the goal. [Neil: hoping] [continues to click and record] (softly) I don't know what color to call that? [Neil: faded wine?] Close enough for me (both chuckle). [continues to click and record] I've run out of color options; I don't even know what to call all of these. Alright, hold on, I'm committed, (Neil laughs). [Neil: salmon or coral?] Coral. Patsy, that's my great aunt and she has a carpet that was that color (the color prior to coral). (Neil laughs)
3	Neil	So, let me pose this question to you: do you think it is possible to find the dark blue that came, that comes after the pink?
4	Danielle	Without knowing the pattern that it is following?
5	Neil	In Room Four.
6	Danielle	(softly) Can't. I'd say that it's possible but not probable in a single person's attention span.
7	Neil	(laughs) So, what if I had someone whose got the world's best attention?
8	Danielle	Eventually, yeah. [Neil: eventually?] At some point.
9	Neil	Now, for the other rooms, I asked you if your method would work no matter state you were in. [Danielle: mmm-hmm (agreement) and color] Would the same method work, so however clicks, say that it takes the person 500 clicks to get back to a blue, a dark blue after a pink. Suppose that we have this salmon now [points to the screen and clicks] and now we have this light pink. Would it take another 500 clicks to get the salmon-light pink combo?
10	Danielle	Probably not. It might, there's a possibility.

11	Neil	So do you think that you could, like does there a general method besides just clicking until you happen to be lucky?
12	Danielle	(laughs) Ummm, how, I don't, there may be, [Neil: maybe] may be a pattern but I think that clicking until you get lucky is probably the easier of the two options [Neil: okay]

When tasked to come up with a new button for Room Four, Danielle brings another aspect of her meaning for randomness to bear: chance. In the clinical interview, Danielle explicitly connected the notion of chance with the notion of randomness, culminating in her indicating that the two were the same thing (see Excerpt 18 and Excerpt 19). This aspect of her meaning for randomness appears in the rule that she gave to her new button as shown in Figure 34.

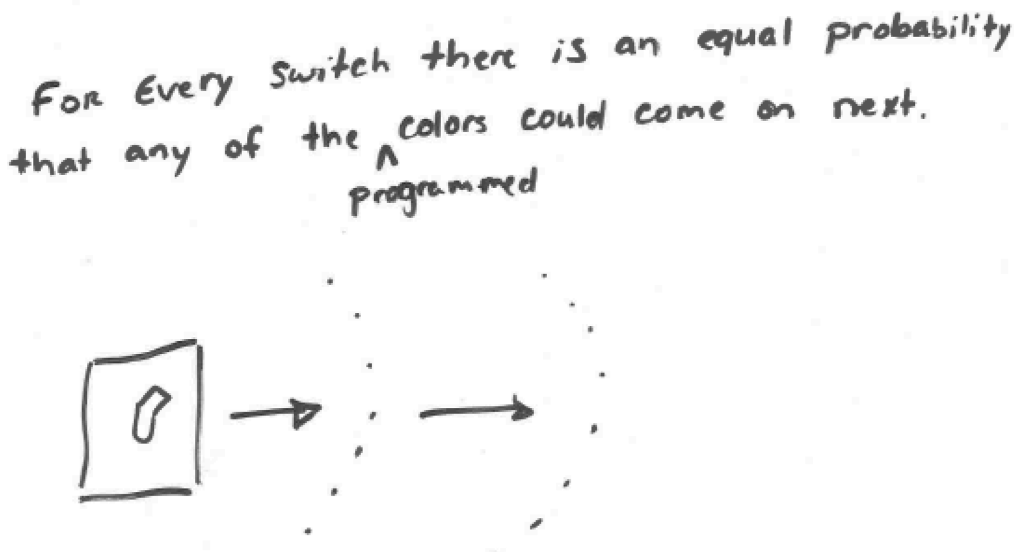


Figure 34. Danielle's new switch for room four

Excerpt 48. Danielle discusses her switch for room four

1	Neil	So, guess what? [Danielle: hmmm] The switch broke. [Danielle: okay] I need you to make me a new switch for Room Four. [Danielle: okay] What would you, what would you have the switch do?
2	Danielle	(long pause) I don't know, pick a new color.
3	Neil	Does that just, so like here (Room Three's sequence) you had the switch move through sequence one at a time, right. [Danielle: no]

		So here (Room Four), what would you have the switch do? You can use English sentences.
4	Danielle	(laughs) So, is this a switch? Can this be a smart switch?
5	Neil	What do you mean by a smart switch?
6	Danielle	Umm, so, of the range of colors that are available for the switch to turn on, there would be like an equal probability of them turning any of them on at any given point when we're clicking.
7	Neil	That is up to you, oh switch designer
8	Danielle	(laughs) Switch designer, hopefully I don't have a job like that in the near future. Umm, so [writes the phrase "For every switch there is an equal probability that any of the programmed colors could come next."] Hopefully I'm using this in the right phrase.
9	Neil	(reading) So for every switch there's an equal probability... that any of the programmed colors could come on the next...?
10	Danielle	On next
11	Neil	On next. So, by that you mean, for the phrase "for every switch" that means every time you press the switch?
12	Danielle	Yeah, every time you hit the switch, or flip the switch, [Neil: okay] or press a button
13	Neil	Okay. Do you think that will work? [Danielle: sure!] So, draw yourself a switch.
14	Danielle	Okay [draws a switch]
15	Neil	And your words there are now the, uh, code behind your button. [Danielle: hmmm-kay] So, if you press your switch, [Danielle: mmm-hmmm] So actually go ahead and press your switch; your switch [Danielle: this switch, okay] I'm going to make you, press your switch. What should happen?
16	Danielle	Umm, [draws an arrow from her switch, and an arc of dots] any of the colors can come up.
17	Neil	[Presses a button for Room 4] So, we got a new color. And if you press the switch again, what should happen?
18	Danielle	Any of the rest of the colors come on, including that one [Neil presses a button for Room Four], well the other one, but okay.
19	Neil	And if you press the switch again?
20	Danielle	Same thing. Okay.

While I did not specifically probe what she mean by equal probability, I suspect that Danielle is adhering to the equiprobability heuristic (Fielding-Wells, 2014; Lecoutre et al., 1990); stated differently, she's operating under the Principle of Ignorance (von Mises, 1981). Given that she cannot discern a pattern, then the button behaves randomly, which would entail every color having the same potential to be called up on any click.

I asked Danielle to create some type of grouping of the Four Rooms. Her first instinct was to order the rooms for least to most complex; One, Two, Three, and then Four. As she continued to play around with groupings, she put Rooms One and Two together as they were both grey scale; she also grouped Rooms One and Three together as they both alternate a color/white with black (i.e., off/on). She noted that she didn't want to put Room Four with any of the other three, prompting me to ask her if she could see a way to put Rooms One, Two, and Three together while Four was a separate group. As shown in Excerpt 49, Danielle views Rooms One through Three as all having patterns whereas Room Four has the hallmark of not allowing for a person to exert to control over the room. That is to say, that Danielle noticed that an individual could choose what color to have lighting/darkening up Rooms One through Three, but not in Room Four.

Excerpt 49. Danielle builds her categories for the processes in the four rooms

1	Neil	Is there a way that you could think about the three of them, Rooms One, Two and Three, being grouped together and Four off by itself?
2	Danielle	Yeah, that would be easier than trying to group this (Room Four) with any one of the other ones, just because, this one (Room Four) is something you can't control and you can't, like, guarantee that you fall in a color. Whereas these guys (Rooms One through Three) have like a distinguished pattern that you can follow, and it the button a certain number of times and you get the outcome that you're looking for.
3	Neil	So maybe that would be a good way to classify them. Of everything that you've said, that's the one you've sounded the most sure about.
4	Danielle	That, yeah.
5	Neil	So let's go ahead and write down our two classifications here.
6	Neil	[Danielle: Okay] on a piece of paper that you come up with. [Danielle: hmm-kay]. You can split the paper in half for now [Danielle: okay] with a, just draw a line, however you want. On one half, will be whatever category you want to call this (gestures to Rooms One through Three), [Danielle: okay] so you get to make up your own name [Danielle writes "Patterned"; Danielle: was I

		supposed to write on this side] and you can come up with whatever name you want for this other category [Danielle writes "randomized"; Danielle: okay.] So, what are your names?
7	Danielle	[Neil: They are...] So, patterned and randomized.
8	Neil	Okay. So, in the Patterned category [Danielle: mmm-hmm] what are you looking for that makes some fit, that makes a process fit? In the patterned category?
9	Danielle	Dependability of [Neil: dependability of what?] Of the process of the outcome, pattern? I guess
10	Neil	Okay, so let's go ahead and write that.
11	Danielle	[Writes: dependability of sequence] Sequence was the word I was looking for
12	Neil	Sequence. So, the dependability of the, of sequence of...?
13	Danielle	Of, switches, clicks, [Neil: clicks, outcomes?] outcomes, yeah [adds of outcomes to her list] Okay. [Neil: okay] Okay, now what's supposed to do
14	Neil	So what about over here in your randomized? What types of things would you look for?
15	Danielle	The inability to predict it.
16	Neil	The inability to what?
17	Danielle	The inability to predict it, to find the sequence, I guess [Neil: okay]
18	Neil	So let's go ahead and write that down [Danielle writes inability to determine sequence] So, the inability to determine a sequence and you've got dependability of sequence which would mean that you would also be able to determine a sequence, right? [Danielle: mmm-hmm, yeah] Okay. Anything else that you would be looking for here in your patterned thing, besides dependability?
19	Danielle	(long pause) I'm not sure.
20	Neil	Do you have questions about the question or...?
21	Danielle	I'm trying to figure out how I'm supposed to answer that.
22	Neil	So think about classifications of plants. [Danielle: okay] Right, and when we, for different types plants, there are all of these different sort of things we look for [Danielle: mmm-hmm] that helps us classify them. That's sort of what you're doing here. [Danielle: like characteristics?] Characteristics. [Danielle: okay] So, characteristics of this particular set and, versus characteristics of this loner.
23	Danielle	(laughs) Alright, so, I guess, depending on which one you get, there's always a pattern between black and whatever other colors there are in there. Umm, so, like this one (Room Three), the characteristics would be alternating black and color, this one (Room Two) the characteristics would be fading from black to white, and this one (Room One) is just black and white.

24 Neil So what, are there any characteristics across all three that make them fit in this patterned category?
25 Danielle Well, all of them have black in them and all of them have a certain number of clicks to get back to the original, whatever you want your original to be. [Neil: okay] So I guess the pattern is only so long
26 Neil Okay. So that could be something. [Danielle writes "pattern is short and defined"] You gave a puzzled look there.
27 Danielle I was trying to figure out how to actually write that down. Okay. So, the pattern is short and easy and defined.
28 Neil Okay. What about over here in randomized?
29 Danielle There is no discernable pattern [Neil: okay] so it is just whatever color happens to be turned on next which sucks for the person living in that room.

Part of Danielle's view of control is tied to her conceiving the processes in Rooms One, Two, and Three as being dependable in terms of outcomes (Lines 9-13). At my prompting, Danielle attempts to come up with additional characteristics for her two categories. At this point, all she adds is that the pattern is short and defined for her Patterned category.

I then present Danielle with the five new processes from the More Processes activity. Rather than labeling the sheets, Danielle added the sheet with each process to her two piles: Patterned (containing Rooms One, Two, and Three) and Randomized (containing Room Four). In discussing why she classified the Baking Soda and Vinegar process (deterministic) in her Patterned category (Excerpt 50), she brings up the notion that the activity can be repeated which she expresses as carrying out the experiment again and again recording that Patterned things are repeatable (Lines 5 and 9). This suggests that Danielle has yet to conceive of repeating the process and repeating outcomes as two separate pieces. Danielle wrestles with this distinction when asked to classify the Ping Pong ball lottery (stochastic) in Excerpt 51.

Excerpt 50. Danielle classifies the baking soda and vinegar process

1	Neil	Which category does this fit in of yours?
2	Danielle	Awww, that kind of question. Are we talking about, what we're actually doing or the reaction that's going to happen after you do this?
3	Neil	Pressing the switch, adding the baking soda to the vinegar, we got an end result, right?
4	Danielle	Ohhhh, okay, I see how you're doing that. Alright, then yeah, it would definitely be like a simple, like (points to her patterned pile)
5	Danielle	We can predict what will happen with this one (baking soda and vinegar) [Neil: but we can predict what happens] Yes, plus cause I've done this and it's fun (Neil laughs). So, yeah, this is predictable, repeatable experiment
6	Neil	So, predictable and repeatable. You've not had anywhere on your category list.
7	Danielle	No, we haven't.
8	Neil	So, should we add repeatable somewhere [Danielle: Probably] Where are you going to add repeatable?
9	Danielle	In the pattern area. [Neil: Okay]. Because if it is randomized, you can't exactly repeat it. [Neil: Okay] [Danielle writes "repeatable" as the third bullet to her Patterned category] Okay.

Excerpt 51. Danielle classifies the lottery

1	Danielle	Well, I can't classify it in this (her Randomization pile) [Neil: Why?] Because I don't know the rest of what they're doing. Like I don't know if the teacher is going to add the ball back to the bag [Neil: they are only drawing out one] Just one? [Neil: just one] Just one. Okay. Hmm. Then in that case, then yes, I would classify it over here (the Randomization pile) because all of the Ping Pong balls are together, and each has an equal probability of being drawn. So, I can't predict which one is going to be drawn.
2	Neil	Suppose that the teacher does this Monday, and Tuesday, and Wednesday, and Thursday and Friday. Did she just repeat her process?
3	Danielle	(Hesitant) Yes.
4	Neil	So, that's repeated though, right (points to her Patterned category characteristics)
5	Danielle	It is repeated, but the outcome isn't...
6	Neil	But all you have on your list repeatable.
7	Danielle	But dependability. (Both laugh) That's not dependable.
8	Neil	What do you mean that's dependable. She is dependable in running this Ping Pong ball lottery.

9	Danielle	But you're dependable in flipping your on switch off and on [Neil: Well] But the outcome isn't reliable. [Neil: ahhhh] Should I be writing that down too?
10	Neil	So maybe you need to uh, to rephrase.
11	Danielle	Rephrase repeatable?
12	Neil	What you rephrase is up to you at the moment.
13	Danielle	(laughs) Alright. [adds outcome behind repeatable in her Patterned list]
14	Neil	You're going to hate me [Danielle: (laughs) Yeah, you're, yeah] Are you ready? [Danielle: yeah] Repeatable outcome. (pointing to the series of three blacks from Room Four) Black.
15	Danielle	Damn it. (both laugh) Alright, umm, there's still not a pattern though [Neil: okay]
16	Neil	So, we can come back to this.

At this point in time, Danielle has firmed up her listing of characteristics for her two categories (see Figure 35). Patterned processes are dependable/predictable; short, defined, and repeatable; and have repeatable outcomes. Randomized processes do not allow for her to determine a sequence/pattern. While she did not list the notion that randomized processes do not allow her to make predictions, she constantly refers to this notion. Danielle places both the Savings Account (deterministic) and Looking at Neil's Shoes (deterministic) processes in her Patterned category. In the case of the Saving Account, she recognized that there was a mathematical formula that would allow her to know exactly how much money would be in the account; this satisfied her condition of dependability. Further she noted that the process of accruing 1% interest was being repeated, even if the same dollar amount was not. While this would seemingly contradict her condition of repeatable outcomes, she noted that the dependability was key. For looking at the color of Neil's shoes, she noted that since my shoes weren't changing between looks, there was a sense of dependability that made that process also fall into the Patterned category.

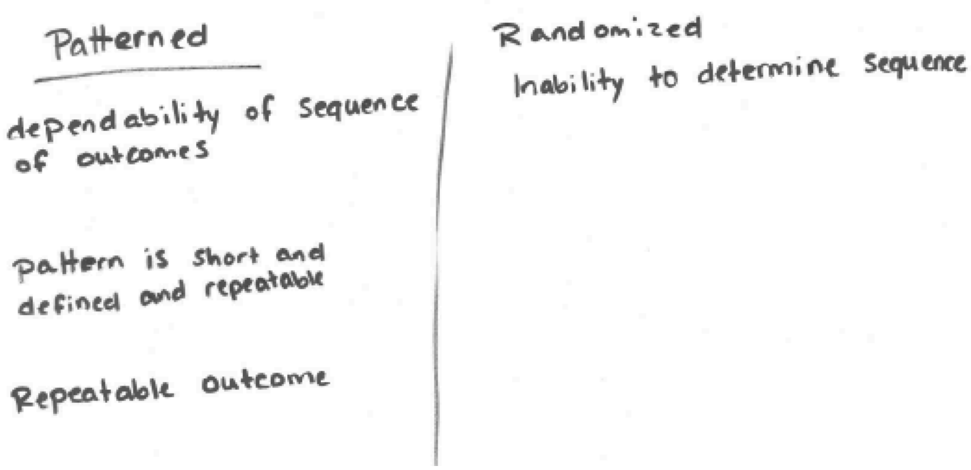


Figure 35. Danielle's two categories

When Danielle encounters the process of recording the total number of red cars going through an intersection (stochastic), she does something very surprising: she makes a distinction between the actions of the car counters and the result of the total number of red cars. In Excerpt 51, Danielle started wrestling with the process being repeated versus the outcomes being repeated. In Excerpt 52, Danielle splits the process out from the result entirely (Lines 1-5). Danielle stayed firmly rooted in her belief that the number of red vehicles could not be predicted. I tested her commitment by bringing up her experiences with one of the rooms in Line 8. She stayed with her placement of this process in her Randomized category, articulating that she could not predict everything and everyone (Lines 7, 11, and 13).

Excerpt 52. Danielle classifies the number of red cars process

1	Danielle	Alright. So, us standing at the same time every day and recording the number, that's predictable but the total number of red vehicles may not be. Because that's it's a busy area.
2	Neil	So. Patterned or Randomized?
3	Danielle	Our actions are patterned, but
4	Neil	But our actions are patterned with the, uhh, Room Four switch. We keep hitting the switch. [Danielle: That is true.] Our teacher's

		actions are patterned, she keeps running this Ping Pong ball lottery [Danielle: that's true] She's going to go broke soon too, right.
5	Danielle	Yeah, I want to be in her class before she goes bankrupt. So, like they are patterned, but the total number of red vehicles that go through the intersection is not.
6	Neil	How do you know?
7	Danielle	Because it is a busy area [Neil: okay] I can't predict everyone else's schedule that day [Neil: okay]
8	Neil	Could you predict what was going to be happening in Room Three before you started exploring Room Three? [Danielle: No.] But once you explored Room Three, could you then predict what was going to happen? [Danielle: yes, that is true] Okay. So, suppose that you start spending your Mondays, 2 to 4...
9	Danielle	I hope not. Alright, so I guess it is random until we find a pattern. But there'd still be like, like there would still be a pretty distinct section of people that just randomly go through Rural and Apache. I can't predict those.
10	Neil	So you're thinking about there's act... there might be two types of people
11	Danielle	Yeah, the people who go through there every day and the people who just happen to be going through there at that exact time. [Neil: gotcha]. Which would throw a kink in the pattern.
12	Neil	So. Which category?
13	Danielle	I'm going to keep it in the randomized since I can't predict everything. [Neil: okay]

Danielle’s images of deterministic and stochastic processes are tied up with her notion of randomness (i.e., can’t find a pattern) and the notion of dependability (i.e. predictability resulting from a pattern). Danielle’s meaning for her Patterned category echoes Bonnie’s meaning: there’s a clear rule and we can determine what comes next. Both students’ meanings for a stochastic process are just the opposite of their deterministic meaning. One way in which the two students differ is that Danielle began splitting the process steps off from the outcomes while Bonnie did not make such a step. Bonnie had moments where she implicitly referenced the idea of replicability; Danielle’s approach the number of red vehicles process highlighted the issues in replicability. Rather than closing the session, I gave Danielle the empty version of the grid shown in

Figure 36. This intervention was intended to help a student flesh out key aspects of a stochastic process: unfixed outcomes (lack of predictability in the short run), separating out the repetition of the process and outcomes (reproducibility vs. replicability) and the nature of the process's rule.

Process	Fixed or Unfixed Outcome?	Type of Repetition: Reproducible or Replicable	Clear or Fuzzy Rule?	Category
Room One	F	Repro	C	• Deterministic
Room Two	F	Repro	C	•
Room Three	F	Repro	C	•
Room Four	U	Repl	F	• Stochastic
Baking Soda + Vinegar	F	Repro	C	•
Labeling Ping Pong Balls	U	Repl	F	•
Savings Account	F	Repro	C	•
Counting Red Vehicles	U	Repl	F	•
Looking at Neil's Shoes	F	Repro	C	•

Figure 36 Danielle's filled in intervention grid

When she first saw the grid (Figure 36), Danielle remarked that she liked the clear or fuzzy rule column. Thus, I elected to capitalize on her focus on that particular column. In Excerpt 53, I ask Danielle what she believes the phrase "clear rule" means. She quickly brings to bear her notions of patterns and prediction (Lines 2-6). This suggests that Danielle took the phrase "clear rule" and associated that with her own image of a process being dependable. She quickly identifies all but one of the deterministic processes as having a clear rule; she struggles a bit with whether or not looking at Neil's shoes and recording the most prominent color has a clear rule or not (Line 12). In Line 13, I took her back through the activity of looking at the shoes in a variety of contexts, when she brings up the idea of not knowing the existence of a clear rule at the start of the process (Line 14). I suspected that her past experiences with light switches, savings accounts, and mixing baking soda and vinegar together provided her meanings to assimilate those processes to, allowing her to fill in gaps. However, her encountering a light switch such as in Room Three (rainbow colors alternating with black) in her past is doubtful. Thus, I used Room Three as an initial site to ask her if we had to know a clear rule before carrying out the process (Line 15). Danielle eventually arrives at the decision that we don't have to know the clear rule prior to carrying out the process.

Excerpt 53. Danielle wrestles with what makes a clear rule

1	Neil	What do you think is meant by the phrase "a clear rule"?
2	Danielle	If, I guess one that you can follow, so like a pattern
3	Neil	So a pattern or something...like, what about Barclays (savings account process)? It's not really a pattern?
4	Danielle	But it will, you can predict it
5	Neil	There's a formula
6	Danielle	Yeah [Neil: right] yeah
7	Neil	We can talk about the interest, or I don't know [Danielle: Or not] (both laugh) [Danielle: I just got out of that] But there's something can do there mathematically [Danielle: yeah] and that would be a clear rule [Danielle: mmm-hmm (agreement)]. So, thinking about all of our processes here, which ones have a clear rule?
8	Danielle	(pauses) I'd say, [Rooms] One, Two, Three
9	Neil	Let's go ahead and put a C
10	Danielle	[marks a C for Rooms One through Three, Baking Soda and Vinegar, and Savings Account] Okay.
11	Neil	I can't quite tell what you're staring at. Because you're like, you're staring at something, like you're, you're staring it down
12	Danielle	(laughs) I was staring down your shoes. [Neil: Okay] Like I, the rule of looking at your shoes is clear, but I can't guarantee, well

now I can because I saw all your shoes, but I can't guarantee what color your shoes are going to be.

13 Neil So, if I'm teaching, what color are my shoes? [Danielle: black] If I'm in my office what color are my shoes? [Danielle: black] If I'm biking what color are my shoes? [Danielle: black, with slightly different colors] But we don't care about the other colors, we only care about the most prominent.

14 Danielle (laughs) Alright, so, I guess...but I didn't know that when I started though.

15 Neil But did you know that Room Three was going to have a clear rule when, before you got started? [Danielle: No] About for Room Two? Did you know that was going to have a clear rule?

16 Danielle I hoped it would [Neil: you hoped] but no.

17 Neil So do you have to know before we begin whether or not there's a clear rule?

18 Danielle Nah, I guess you're right.

After she labels all of the processes that she believes have clear rules, Danielle tackles the remaining three; Room Four, Ping Pong ball lottery, and counting red vehicles. Just as she connected a clear rule to having a pattern, she made fuzzy rules as having no pattern (Excerpt 54). For Danielle, being able to establish a pattern, even by talking to every driver (Line 8), is paramount for determining whether or not there is a clear rule. Her being able to find a pattern is what will ensure that she sees that three hallmarks she came up with (Figure 35); especially, the defined, repeatable pattern and the dependability of the sequence of outcomes.

Excerpt 54. Danielle describes fuzzy rules

1 Neil So then that leaves Room Four, the [Danielle: Ping Pong] Ping Pong ball,

2 Danielle and the red vehicles

3 Neil as being fuzzy. [Danielle: mm-hmm (agreement)] What makes something fuzzy?

4 Danielle No pattern

5 Neil No Pattern. Is there a mathematical formula, like the Barclays, that we can do for the, uhh, Ping Pong ball lottery?

6 Danielle There is. [Neil: There is?] I don't remember what it was, but yeah there is. [Neil: shakes head no] There isn't? [Neil: nope] I thought

you could...ahh, I guess you can calculate the likelihood [Neil: you can calculate probabilities and likelihoods] Yeah, but you can’t guarantee, okay.

7 Neil And is, is there a clear rule for the, uhh, intersection?

8 Danielle Not unless I talk to every single driver.

9 Neil But even then, what about Sally from out of town who rents a car that happens to be red?

10 Danielle Ahhh, no. I wouldn't be able to that one.

11 Neil So there's no clear rule there? [Danielle: mm-nuh] But is there, what about Room Four? Is there a clear rule? [Danielle: No] But is there enough of something that we can carry out the process?

12 Danielle Yeah.

Excerpt 55. Danielle and fixed/unfixed outcomes

1 Neil So, an outcome is what to you?

2 Danielle An outcome?

3 Neil The result of doing something? [Danielle: yeah] So all of these are processes, right? [Danielle: okay] Room One had the process of pressing the switch [Danielle: mmm-hmmm] we got a room color [Danielle: yeah]. So, what does it mean for something to be fixed?

4 Danielle Guaranteed. The patterned (points to her Patterned process pile)

5 Neil So which things have fixed outcomes [Danielle: those (points to her patterned pile)] Which things have unfixed outcomes [Danielle: those (points to her randomized pile)] So you can fill in that column, right?

6 Danielle Yeah, super easy. So, a fixed outcome. I keep trying to put on the (refers to pen cap) [write F for Rooms One, Two, Three, Baking Soda and Vinegar, Savings Account, and Looking at Neil's shoes] Okay.

7 Neil And for the unfixed. [Danielle: would be the other three] Now notice what lines up. Clear rules and fixed outcomes; fuzzy rules and unfixed outcomes [Danielle: mmm-hmm (agreement)]. So, notice that we can thing about these two things now as being linked together [Danielle: mmm-hmmm]

From this point, I moved Danielle on to the notion of fixed and unfixed outcomes (Excerpt 55). While Danielle seemed unsure of my initial question (Line 1), she did appear to agree with viewing an outcome as the result of doing something. She takes the idea of a fixed outcome and refers to that outcome as being guaranteed. This is in line with her viewing a deterministic process as having a dependable pattern. Danielle quickly labels all of the processes in her patterned pile as having fixed outcomes while the processes in her randomized pile get the unfixed label.

While she was initial working through the Four Rooms and More Processes activities, Danielle used the word “repeatable”, eventually leading her to describe both the pattern as being repeatable and having repeatable outcomes. In Line 1 of Excerpt 56, I highlighted that she had gotten repeated outcomes in a situation that she did not view as being dependable (Room Four) in attempt to help her make a distinction between repeating a process and repeating the outcomes. Danielle initially describes replication drawing from what she has learned from her sciences courses: replication is repeating a study and getting the same results (Line 8). While she wants the results to be exactly the same, she eventually lets that aspect go for consistent results and lets reproducible take on getting the exact same results (Lines 10-12, 18-22).

Excerpt 56. Danielle confronts replicable and reproducible

1	Neil	You've used the word repeatable [Danielle: yeah] but I've point out that we can repeat all of the processes, but then talked about outcomes being repeated. Then I pointed out that then we got repeated outcomes here (points to her list of colors from Room Four) [Danielle: mmm-hmm]. You've also talked about dependability, so like even though we had black repeated (in Room Four), the sequence wasn't dependable, there wasn't a pattern [Danielle: yeah] Now notice that nowhere in here (the intervention grid) I've really used the word pattern.
2	Danielle	Nah, I actually like replicable a lot better.
3	Neil	Okay. So, notice that there are two choices here: reproducible and replicable. [Danielle: mmm-hmmm] Have you ever heard those two terms before? [Danielle: yes] Where have you heard those terms before?
4	Danielle	Typically in my biology classes
5	Neil	Ahh, how do they use those words in biology?
6	Danielle	Umm, so if someone does a study of something, it has to be replicable.
7	Neil	What does that mean?

8	Danielle	It means that if someone else goes out there and does the exact same thing, they should get the exact same outcome.
9	Neil	Exact same?
10	Danielle	Well, baring human err and plans...[Neil: exact same thing?] No.
11	Neil	Not exact but at least compatible, right [Danielle: yeah.] Right, so if you go and do a study. And I go out and take your methods and do the same study [Danielle: mmm-hmm] but say three years later, [side conversation on publishing time delay] and so if we replicate something, we're looking for what we say are consistent results [Danielle: mm-kay] but they don't have to be the same.
12	Danielle	Ahhh, 'kay.
13	Neil	What about reproducible? Where have you, might have heard that term before?
14	Danielle	The longer I stare at the word, the less it sounds familiar. So, you said that replicable is...
15	Neil	Repeat the methods and we get consistent results, or we hope we get consistent results
16	Danielle	Close enough to consistent, and then reproducible, dang
17	Neil	Have you ever heard of reproduction? Where have you heard the word reproduction before besides [Danielle: biology] biological reproduction?
18	Danielle	(laughs) Ummm, probably like, I don't know in the art industry I guess. [Neil: art?]
19	Neil	They make lots reproductions of famous reproductions of famous paintings. [Danielle: mm-hmm] There's a famous company who has made their living off of reproducible-ness. They're called Xerox.
20	Danielle	Oh! So, like, okay, so a copy.
21	Neil	So copy. [Danielle: mm-hmm] So something that is reproducible not only can be repeated but we get...
22	Danielle	The exact same thing.
23	Neil	The exact same thing. [Danielle: okay] So in this list of processes, which ones are reproducible and which ones are replicable?
24	Danielle	Okay. So then, basically, they follow the same.
25	Neil	So why do you think they would follow the same?
26	Danielle	These three (points to Room Four, Ping Pong ball lottery, Red Vehicles), I mean you follow the same process but you're not going to get the same outcome [Neil: ahh] So then they wouldn't be reproducible, they would have to replicable [Neil: Okay]
27	Neil	So let's go ahead and fill those in. Sorry, there’s a double r.
28	Danielle	No worries [writes Repl for the three stochastic processes]
29	Neil	Okay. So, the other ones would be?
30	Danielle	Reproducible

Excerpt 57. Danielle is introduced to the formal names deterministic and stochastic

1	Neil	Go ahead and use blue [Danielle: to give the black a rest] To give the black a rest but also to, put a blue dot in all of the ones that you think are actually going to be in the same category.
2	Danielle	[places a dot in Rooms One, Two, Three, Baking Soda and Vinegar, Savings Account, and Looking at Neil's shoes]
3	Neil	So notice that you've got the same ones that you sort of have to begin, right? [Danielle: mmm-hmm] Everything you put a blue dot in, is something that we call a deterministic process.
4	Danielle	I've heard of that word. Should I be writing that down?
5	Neil	Yeah; if you want to write it in all, you can; if you just want to write it in one, that's fine. So, when we say that something is deterministic, what do we mean by that phrase?
6	Danielle	You can reproduce it and it has a fixed outcome. [Neil: and?] And it's clearly defined through your rules.
7	Neil	Alright. So, in the other ones, the three that don't have blue dots, we call these stochastic.
8	Danielle	There's that word. [Adds red dots to Room Four, Ping Pong ball lottery, and Red Vehicles]
9	Neil	S T O [Danielle: thanks] C H A S T I C
10	Danielle	I thought it had an I in it for some reason.
11	Neil	You might be thinking of stoichiometry.
12	Danielle	That's what I was thinking of, okay.
13	Neil	Stochastic processes have fuzzy rules, unfixed outcomes, and are replicable. [Danielle: mmm-kay]

Now that she has filed in the three columns, I introduced the formal names of deterministic and stochastic. She connects the three columns to her past experiences with the term “deterministic”, especially the ideas of fixed outcomes, reproducible, and having a clear rule (Line 6 of Excerpt 57). Danielle appeared to have taken the ideas of fixed outcomes and a clear rule as linking to her own patterned category for establishing a dependable pattern. The lack of those two aspects (i.e., having a fuzzy rule and unfixed outcomes) she then linked with her randomized category. At this moment, her image of reproducible as making copies fits within the dependableness of patterned processes.

Teaching Experiment Session Two

Danielle's second session focused on her exploring the Sequences applet while using a grid based upon the intervention that she saw at the end of the first session of the teaching experiment. Her filled out grid for this session appears in Figure 37. As a whole, Danielle classified the fourteen process in a normatively correct way.

Before engaging with the Sequences Applet, I asked Danielle to recap of the prior sessions (Excerpt 58). I should note here that approximately two weeks had passed between the prior session the present session. After looking at the blank intervention sheet (Figure 37), Danielle was able to recall the terms deterministic and stochastic and then provide explanations of those terms (lines 2–6 of Excerpt 58). She highlighted that fixed outcomes always happened while unfixed outcomes could the result of a pattern that an individual might not understand or see (lines 9–11). As she explained the ideas of unfixed/fixed outcomes, reproducible/replicable, and clear/fuzzy rules, she often brought up the situations she had worked through in the prior session. Primarily, she brought up Room One as the fixed outcomes example for fixed outcomes pointing out that there would always be a black-white alternating pattern. When a person couldn't catch onto a pattern in the outcomes, Danielle described this as unfixed outcomes (lines 10–12). She used the savings account situation as her example for both reproducible and clear rule. Danielle conveyed that reproducibility meant that a person could repeat the activity again and again, getting the same results (lines 14–18). Since we could establish a formula for the savings account, that meant that there was a clear rule that we could use (line 24).

Process	Fixed or Unfixed Outcome?	Type of Repetition: Reproducible or Replicable	Clear or Fuzzy Rule?	Category
Zero	F	Repro	clear	Deterministic
One	F	Repro	Cl.	Det.
Two	~~F~~ U	~~Repro~~ Repl.	~~C~~ F	~~Det~~ Stochastic
Three	F	Repro	C	Deterministic
Four	F	Repro	C	Det
Five	u	Repl.	F	St.
Six	u	Repl.	F	St.
Seven	F/u	Repl.	C	Det.
Eight	F	Repro.	c	Det.
Nine	F	Repro.	C	Det.
Ten	F	Repro.	C	Det.
Eleven	u	Repl.	F	Sto.
Twelve	F	Repro.	F	Det.
Thirteen	u	Repl.	~~C~~ F	Sto.

Figure 37. Danielle's filled out grid for the sequence's applet

Excerpt 58. Danielle's recap of deterministic and stochastic processes

1	Neil	So what were the two types?
2	Danielle	Deterministic and stochastic.
3	Neil	Yep. Do you remember what made a process a deterministic process?
4	Danielle	It was the rules, correct? It was mostly based on the rules or was it based on all three? Yeah, it was based on all three. [Neil: okay] So, deterministic was clear rule, umm, reproducible, and then with a fixed outcome.
5	Neil	Okay. And the other category was?

6	Danielle	Stochastic, which had fuzzy rules, and it was replaceable [misspeak?], and it was unfixed.
7	Neil	So, just as a recap, what is a fixed versus an unfixed outcome? Do you remember?
8	Danielle	Oh, vaguely. Umm, a fixed outcome was...it was always like, the outcome was always guaranteed. So, we'd, black white black white, [Neil: okay]
9	Neil	And the unfixed?
10	Danielle	The stochastic was the one where, oh man, it has been like two weeks, ummm, basically, you couldn't judge it, like you couldn't, ummm, catch on for the lack of a better word.
11	Neil	Okay. And what would you be catching on to?
12	Danielle	The pattern behind
13	Neil	How about reproducible versus replicable?
14	Danielle	So, if I remember correctly, reproducible was you can guarantee what would happen every single time, so after black there would be a white tile. And then replicable would be...words, ummm, replicable would be like you can get there eventually. I was trying to think of the last one, Room Four.
15	Neil	So, here you have the savings account as being marked as reproducible. [Danielle: mm-hmm] Would that fit with your description prior, that you just gave?
16	Danielle	(coughs) Excuse me
17	Neil	Do you remember we were putting the same amount of money [Danielle: yeah] at the same interest rate.
18	Danielle	Yeah, so reproducible means that you can do it all the way through from the beginning.
19	Neil	Here with Ping Pong balls we have replicable.
20	Danielle	So you can do the process, but the end result isn't the same [Neil: okay] Okay.
21	Neil	So reproducible the process and the end result are the same? [Danielle: Yeah] And replicable the process is the same but not the end result?
22	Danielle	Yeah, there we go, that's much better way then I was trying to explain it.
23	Neil	And what do we mean by a clear or fuzzy rule?
24	Danielle	So, now that I actually read these on the side, I'm going to stick with these two. Ummm, so like a clear rule would be, the bank, the saving account has like a formula you could follow. So, it's a formula you follow every single time and then a fuzzy rule would be like with the Ping Pong balls. You pull one out, but you don't pull any particular one out. So, there's like always a level of variability to it. So, it's not clear which one will come out next. [Neil: okay] as opposed to the other one.

For the unfixed outcomes, fuzzy rule, and replicability, Danielle used both Room Four and the Ping Pong Ball lottery as her examples. Her conveyed image of replicability centers on the ability to imagine carrying out the process over and over again, but you won't necessarily see the outcomes in the same way (Excerpt 58, lines 14, 19–22). However, Danielle's image entails an anticipation that if she were to repeat a replicable process long enough, she could see particular event of interest (black followed by white; line 14). I take her statements to indicate that Danielle is starting to develop a disposition to think in the long-run; that is, imagine processes running infinitely many times rather than focusing on immediate outcomes.

After her recap, I had Danielle work with the Sequences Applet. I presented Process One to Danielle. After seeing the first three terms, Danielle was starting to look for a pattern. After the next three terms, she felt confident that she understood the pattern (Excerpt 59, lines 1–3). Danielle believed that the pattern was that each number was repeated that number of times (line 7). However, when that pattern was broken, Danielle continued to search for a pattern that would allow for the single four (lines 7–15).

Excerpt 59. Danielle looks for a pattern in process one

1	Danielle	(Neil clicks, 1) Okay. (Neil clicks, 2) Kay. (Neil clicks, 2). Alright.
2	Neil	I see you squinting at the screen. (Danielle laughs) What are you thinking at the moment?
3	Danielle	I'm trying to figure out how it is going to trick me this time. [Neil: kay. Are you ready?] Yeah. (Neil clicks, 3). Alright. (Neil clicks, 3) I think I have a handle on it. Let's do one more (Neil clicks, 3). Okay, I think I know the pattern.
4	Neil	Are you sure?
5	Danielle	I'm never sure, but I'm pretty sure.
6	Neil	Shall we try another? What do you think the next one will be?
7	Danielle	I think that the next one will be four. (Neil clicks, 4). [Neil: And what do you think the next one will be?] I think three will be fours. (Neil clicks, 5) Dang it! (laughs) Okay. [Neil: So, one more?] Sure.

(Neil clicks, 5) Alright. (Neil clicks, 5). That screwed up my plans again, dang it.

8 Neil How'd that screw up your plans?

9 Danielle Alright, I was trying to find a pattern, so I figured since that was a four and that was five, then maybe it repeated the pattern back to one. So, if that was going to be the case, then there'd be two fives and then three sixes.

10 Neil Okay. But how many fives are there?

11 Danielle There are three fives, which means that did not work.

12 Neil Okay. So, what do you think will happen next?

13 Danielle I have no idea at this point [Neil: shall we find out?] Yeah. (Neil clicks, 5) Umm, kay. [Neil: Again?] yeah. (Neil clicks, 5)

14 Neil So what are you thinking?

15 Danielle I'm trying to figure out how the four plays into all of this. Because it was going well, 1 one, 2 twos, 3 threes, and then four happened and then that screwed me up (laughs). And now there's 5 fives, so far. [Neil: mmm-hmm] So I'm just trying to figure out where the four plays into all of this.

16 Neil What do you think you'll get next? [Danielle: Hopefully a six.] Shall we find out? [Danielle: sure] (Neil clicks, 6)

17 Danielle Okay. Hmm.

18 Neil So what do you think you're going to get after the six?

19 Danielle I think a seven.

20 Neil Why do you think a seven?

21 Danielle Because...alright, so there's 3 threes, a four, and then 5 fives. So maybe there's a pattern then the next two will be sevens and then there will be 8 eights after that point. [Neil: shall we find out?] Sure. (Neil clicks, 7) Okay. (Neil clicks, 7) So, if that pattern is true, then the next one should be an eight. (Neil clicks, 7) Dang it.

22 Neil Shall we try another one [Danielle: yeah] (Neil clicks, 7)

23 Danielle So maybe the next three will be sevens and then there will be another eight.

24 Neil So let's try this. (click) Seven. (click) Seven. (click) Seven. And then we should have an eight? [Danielle: yeah] (click, 8)

25 Danielle Okay.

26 Neil So then what should we have?

27 Danielle A nnn....a nine. (Neil clicks, 9) Okay, there should be like 8 nines after that point. (Neil clicks, 10). Gosh darn it. Okay.

28 Neil So what will come next?

29 Danielle (long pause) I'm not sure. (Neil clicks, 11) So now it goes back to normal.

30 Neil Goes back to normal?

31 Danielle Maybe. Eight, nine, ten, well seven, eight, nine, ten, maybe. So, it's not really a predictable pattern that I can see.

Danielle's search for a pattern in Process One involved some aspects of a positional pattern (Excerpt 59, lines 16–21). At first, I thought that Danielle was imagining that the value repetition ran in sets of three: the first term in a set does not get repeated, the following two terms would. This would give Danielle the ability to explain the four. However, she did not imagine the six being repeated (lines 18-19). Rather, she appears to have reset her set of three with the six. As I continued to add more terms, Danielle continued to look for some kind of pattern.

As Danielle searched for a pattern, she did not think about the sheet nor the three aspects I had her recap. I took her to mean in line 31 of Excerpt 59 that she did not see something that she could take as a clear rule. Thus, after filling the applet's grid, I directed her attention to the intervention sheet (Excerpt 60). There are a couple of interesting points that happened at this time. First, when discussing which kind of repetition (reproducible vs. replicable), Danielle focuses on just the outcomes being repeated (line 4). With a couple of questions about the relationship between the repeated numbers, Danielle notes that the repeated numbers are all prime (lines 5–8). Second, when I refocused her on type of repetition, Danielle was able to shift away from the outcomes being repeated (or not) to the process. When she watched the process being repeated and observed getting the same values in the same order, she was able to bring together the aspects of primes with prior pattern to establish a rule (line 18).

Excerpt 60. Danielle classifies process one

1	Neil	Shall we fill in the whole grid? [Danielle: yeah]
2	Danielle	Yeah, that one's not a clear one.
3	Neil	So we don't have a clear rule at the moment [Danielle: no, that I know of] Do you think we're fixed or unfixed? [Danielle: I think it's unfixed] Okay. What type of repetition do we have here?

4 Danielle Well, at first it started off pretty good with like the number of each individual number following the numerical value. And then four happened and that screwed everything up. So, two, three, five, seven, eleven, and I don't know about thirteen, but those all follow the same

5 Neil So keep in mind that we are limited to only 40 terms. [Danielle: okay, so it stops after that point?] Yep. So, we don't have any access beyond the 40th trial [Danielle: okay] But, you, you listed off two, three, five, seven, eleven, and thirteen [Danielle: yes] is there anything special about those numbers?

6 Danielle The number of tiles they take up is there numerical value [Neil: mm-hmm, is there anything more, else special about them?] They're odd numbers, well, except for two.

7 Neil Nine's an odd number.

8 Danielle Hmmm, they're...[Neil: they're what] they're primes. [Neil: they're primes] Dang. I didn't think about prime numbers.

9 Neil So let's come back though to this idea of the type of repetition thought. That's, this is about the process, right, not the pattern. [Danielle: mm-hmm, right] So what would we need to do to help you to decide whether this is reproducible or replicable?

10 Danielle Do it again. [Neil: Alright, shall we do this again?] Sure.

11 Neil So we'll start back at the first outcome. [Danielle: mmm-hmm] (Neil clicks through the process)

12 Danielle Okay. Yeah, so it is following the same rule.

13 Neil So are we reproducible or replicable?

14 Danielle Reproducible [Neil: okay]

15 Neil So do you think we're fixed outcomes or unfixed?

16 Danielle Now that I know the pattern behind it, I think it is a fixed outcome.

17 Neil Do you know the pattern? [Danielle: yes] So, fixed, okay let's go ahead and mark that. And if you know the pattern, would we have a clear or a fuzzy rule? [Danielle: Clear.] So, what's the pattern?

18 Danielle Deterministic. No, the pattern, not the category. The pattern is, umm, so every number is repeated once, but prime numbers are repeated their numerical value. [Neil: Okay. Does that work as a rule?] Don't see why it wouldn't.

Excerpt 61. Danielle pattern hunts with process two

1 Neil First trail (clicks, C) [Danielle: okay] Second trial. (clicks, A) [Danielle: okay] (clicks, C) [Danielle: okay] (clicks, C) (clicks, C). [Danielle: Alright.] So, what's going through your mind?

2 Danielle Well, a simple pattern so far, kind of crossed a list because no there's a whole bunch of C's. [Neil: okay] But so far I just see an A. So, there may be a pattern. (Neil clicks, G) Well, that screwed that

one up. (laughs) Aright. You've got this down to an art form (Neil laughs). Alright (Neil clicks, G). Okay. (Neil clicks, A) Alright,

Following Process One, I took Danielle through Process Two (DNA). Immediately, Danielle reverted to hunting for a pattern in the values (Excerpt 61). Curious, I asked Danielle if she was thinking about the attribute on the intervention sheet (Excerpt 62). For Danielle, looking for a pattern was the key way that she could find a rule. If she could find a rule to follow (a clear rule), then she could classify the process. The other attributes took a back seat to this aspect for her.

Excerpt 62. Danielle describes how patterns fit with the sheet

1	Neil	Still looking for a pattern? [Danielle: yep] What about these three attributes here? [Danielle: So far?] hmm-mm (agreement) Are you thinking about them any, or just pattern?
2	Danielle	Well, I guess pattern plays into it for me. [Neil: okay] So if I can find a pattern it means that it automatically, there has to be rule that it follows [Neil: okay], if I can figure out the pattern.
3	Neil	Was there actually a pattern in that last one (Process One) or just a rule? A pattern is something that sort of repeats
4	Danielle	A rule. I guess that's true, but the rule pattern makes a pattern after a while. I guess my definition is slightly different. (laughs)
5	Neil	That's okay. We're just, I'm just trying to figure out what your meaning for the word pattern is.
6	Danielle	Alright, so, I guess a pattern is something that has a rule that I can follow. [Neil: okay] And so far, this one doesn't have one. [Neil: okay] (Neil clicks, T) Even less of one now.

I shifted Danielle’s focus from looking for a pattern to trying to place the letters into a context. After dropping a hint about biology and making a double helix, Danielle made the connection between the letters and a sequence of DNA nucleotides. Given her proclivity to look for a pattern, I decided to frame the DNA sequence with in a context where investigating whether the process was reproducible or replicable could create perturbation in her thinking. To do this, I framed the process as if the process lists off a single person’s DNA sequence (Excerpt 63). With this initial image, Danielle stated that the Process Two would have fixed outcomes, be reproducible, and have a clear rule (lines 2–10).

Excerpt 63. Danielle’s first classification of process two (single person’s DNA)

1	Neil	If this is a single person's DNA, and we're just reading off, do you think that is going to be fixed or unfixed in terms of what the next nucleotide is for this person?
2	Danielle	It would be fixed.
3	Neil	Fixed. Would the process of looking at the nucleotides be reproducible or replicable?
4	Danielle	Looking at the nucleotides, would be reproducible.
5	Neil	Okay. So, would we have a clear or fuzzy rule?
6	Danielle	It would be clear in that case.
7	Neil	Let's go ahead and mark that in. [Danielle: okay] So if this is a single person's DNA, we would have fixed outcomes, [Danielle: mmm-hmm] If we're reproducible, right? [Danielle: mmm-hmmm]
8	Danielle	And then clear rule, if you're just looking at it and trying to...(trails off)
9	Neil	So then what type of category would this be?
10	Danielle	It would be deterministic. [Neil: okay]

Given that I had set Danielle up for a perturbation, I asked her how we could check whether Process Two as actually reproducible. As I re-ran Process Two a term at a time, Danielle ran into an unexpected “T” (line 4 of Excerpt 64). Of the three possible explanations that Danielle comes up with (human error, DNA changing, or not reproducible), Danielle focuses in on reproducibility/replicability (lines 7–10).

Excerpt 64. Danielle wrestles process two

1	Neil	How could we check if they actually are reproducible?
2	Danielle	Sequence it again. [Neil: okay] (Neil clears the applet) [Neil: Are you ready?]
3		[Neil provides the first several terms from the first time through Process Two: C A C C C C]
4	Danielle	Oh, that's right. (Neil clicks, C) Okay. (Neil clicks, A) (Neil clicks, T) Hmmm
5	Neil	What's the hmmm?

6 Danielle So, if they started from the same starting point, they messed up somewhere. Or the DNA changed. Or it's not reproducible and it's replicable.
7 Neil So, fill in the whole shebang here [Danielle: I should have written that down, shouldn't have I] So, do you think that this is now still reproducible? Or maybe now not reproducible?
8 Danielle No, I don't think that it's reproducible anymore because you didn't get the exact same result.
9 Neil So maybe we're replicable. [Danielle: yeah] Why would we be replicable?
10 Danielle Because you could still sequence the DNA in the same way, the process is the same but your outcomes different.

With her image of process two now being replicable instead of reproducible, Danielle changes her original answers and reclassifies the process (Excerpt 65).

Excerpt 65. Danielle re-classifies process two

1 Danielle So, (looks at her answers) it means all my answers are different.
2 Neil Why do you think that means all of your answers are different?
3 Danielle Because if it is not, if it is not reproducible exactly, then your rule can't be as clear as you would like it to be. [Neil: ahh] And that would mean, that by definition your outcome wouldn't be the same, so you'd be unfixed. [Neil: okay]
4 Neil Cause that's also what replicable is, right? [Danielle: mmm-hmm (agreement)] We can carry out the process, but we don't, by your own admission then the end results aren't the same. [Danielle: Yeah] So let's do that; let's go ahead and fix the rest. So, what would that mean for category?
5 Danielle It would be stochastic.
6 Neil So, I want you to think back on what you did with these three (Processes Zero through Two). [Danielle: Okay] With the first two, you were looking for a pattern. So, in the third you were looking for a pattern. [Danielle: mmm-hmmm (agreement)] But then I had you think about the type of repetition. [Danielle: okay] Notice that by investigating the type of repetition, rather than pattern hunting (Danielle chuckles), you were able to better identify. [Danielle: That's true, yes.] Okay, I want you to keep that mind. [Danielle: Alright].

After she finished re-classifying Process Two, I asked Danielle to think back through her activity in the first three processes. When working with Process Three (deterministic, $2n + 1$), Danielle quickly saw a pattern that worked. However, after she had me fill the grid, then had me clear the applet and re-run Process Three. Upon seeing the process work identically as the time before, she classified the process as being reproducible and quickly finished classifying Process Three. For Process Four (deterministic, $n^2 / 100$), Danielle acted in a similar manner: she looked for a pattern (initially seeing the perfect squares, struggled for a bit on the dividing by 100) and then checked the process by having me reset the applet and re-running the process. At this point, Danielle's first instinct is to still look for a pattern, however, she is now using the notions of reproducibility/replicability to serve as a check on a pattern that she believes to be holding. With Process Five (stochastic, standard normal distribution), Danielle encountered a situation where she could not discern a pattern within the first several terms as she could with Processes Three and Four (Excerpt 66). As Danielle looks for a pattern, she is pre-disposed to try to find a formula like she did in Process Four. However, the appearance of a negative number (line 4) throws her for a loop. When looking at the filled in applet (line 5), Danielle can't find a pattern. Given what that she needed a pattern first before she checked reproducibility/replication with the prior processes, I prompted her (line 6) to recall our discussion after Process Two. Danielle expresses that she hopes that the values will not be repeated in the same order as that would indicate that the process is reproducible and would therefore have a rule (lines 9–15). After stating the Process Five is replicable, Danielle quickly categorizes the process as having unfixed outcomes and a fuzzy rule before declaring the process stochastic (lines 17–25).

Excerpt 66. Danielle classifies process five

1	Neil	So what are you thinking about?
2	Danielle	The last one was, there was a formula you could figure out after a while. I was trying to see if maybe there's something to work with these guys, but ...(Neil clicks, 0.61)...Hmm, I still don't see one though [Neil: okay] (Neil clicks, 0.2). Hmmm.
3	Neil	So what things are you thinking about as you're looking for whatever you're looking for?
4	Danielle	Alright, well, it starts off with a number greater than one, and then everyone after is less than one, but they aren't in any, particular direction. [Neil: okay, what do you mean by that?] And so, umm, it's not like a 1.22 and then 0.87 then 0.73, like they are switched. It's not a continuously smaller [Neil: so, we went down then up and then down] yeah. [Neil: and then down again] Yeah. [Neil: Okay. Rather than down up down up down up down up, or all down] Or just continuously, yeah, yeah. [Neil: okay] So, (Neil clicks, –0.14) Well now not all of them positive now. At least they had that going for them…
5	Danielle	…I still can't see like any correlation between them [Neil: okay] (Neil clicks, 1.2) Yeah, I don't see a pattern to any of these guys. [Neil: Shall we fill the gird] Yeah. A lot of them are similar, and a couple of them repeat.
6	Neil	Are you still looking for pattern? [Danielle: Trying to, but I don't see one] So you remember our conversation we had a little bit ago? [Danielle: Yes, I know (laughs)] So, what should we try to do?
7	Danielle	We should try to do it again. And hopefully I'll remember the first five or so (Neil points to scratch paper). [Neil: Do you have paper?] I do. Oh, hey! Look at that; that would be useful. (Danielle records several values.) Okay.
8	Neil	Alright. So, what are you hoping for when I press this run process once button?
9	Danielle	I'm hoping that they don't repeat.
10	Neil	And what's the thing that's being repeated?
11	Danielle	The numbers, like I hope they aren't repeated in the exact same order.
12	Neil	So, we shouldn't get a 1.22 in the first position? [Danielle: yes] What if you get a 1.22 in the second position? [Danielle: then I'd be okay with that] Okay. But if, what if you get a 1.22 in the first position?
13	Danielle	If I, I'm hoping that I don't get these guys in order as they are in the first six. [Neil: okay] Because that just, that would mean that they aren't not replicable, or not reproducible but replicable.
14	Neil	And if you do get the first, those six in order?

15 Danielle Then it means that it is reproducible, I just don't know the rule that it's following. [Neil: Okay. Shall we find out?] (Neil clicks, –0.7) Yay! [Neil: you're cheering] I am cheering (laughs) [Neil: shall we do the next one?] Sure. (Neil clicks,–0.02) Okay. (Neil clicks, –1.55) Yeah, so it's not reproducible. [Neil: Are you sure?] Pretty sure. [Neil: Positive?] Yes [Neil: Shall we fill?] Yeah. Mostly these are negative this time, that's cool. So, it's not reproducible, which means it's replicable.

16 Neil Okay. What does that mean for the type of repetition?

17 Danielle Which means it's, the process is replicable, but the results are not going to be the same. (Marks the sheet)

18 Neil So if we're replicable, what does that mean for your outcomes?

19 Danielle Definitely unfixed.

20 Neil Okay. What does that mean for your rule?

21 Danielle It's fuzzy as far as I know.

22 Neil Okay. Definitely not clear?

23 Danielle No, definitely not clear.

24 Neil So what does that means in terms of your process?

25 Danielle That it would be stochastic not deterministic.

Figure 38. Danielle's record of process six values

The success that Danielle had with Process Five carried over to Process Six (stochastic, standard die roll). While I believe she still looked for a pattern to begin with, Danielle took steps that she had not previously done as quickly as she did: she recorded the first set of values (see line 2 of Excerpt 67 and Figure 38) and she prompted me to fill then clear and redo Process Six (lines 2–4).

Excerpt 67. Danielle classifies process six

1 Neil Process six. Are you ready? [Danielle: yes] (Neil clicks, 3) [Danielle: okay] (Neil clicks, 2) [Danielle: Okay.] (Neil clicks, 6)

[Danielle: mmm-kay] You squinted your eye at that six. (Danielle laughs) What did the six do to you?

2 Danielle It was there, it was, it existed. [Neil: okay] Okay (Neil clicks, 1) Okay, so 3 2 6 1, right [Neil: mmm-hmm (agreement)] (Danielle writes these values down) Okay. (Neil clicks, 1) Alright. [Neil: do you think you know what's going on?] Not a clue but I can check it. [Neil: You can check it? What are you going to check?] So, if you fill it in and then redo it. [Neil: So, fill in.] Okay.

3 Neil So you want me to clear and re-run the process? [Danielle: Yes.] Alright. (Neil clicks, 3)

4 Danielle Okay. (Neil clicks, 6) [Neil: shall we go again?] Yeah. (Neil clicks, 3) Okay. (Neil clicks, 5) Alright. (Neil clicks, 1)

5 Neil Shall we fill in? [Danielle: yeah] So what are you thinking about this process?

6 Danielle Hmm, that it's, they aren't the same, so it's not reproducible. [Neil: what's not the same?] The end results, the values that actually appear [Neil: okay] So our results aren't the same [Neil: so, what type of repetition?] It's replicable. [Neil: okay. Are you sure?] No. [Neil: How can you convince yourself even more so?] Do it again. [Neil: Do it again? Okay. Are you ready?] Yep. (Neil clicks, 4) Okay (Neil clicks, 1) Yeah, so I'm sure (Neil fills the board). Cool. So, it is replicable again [Neil: okay] Unfixed outcome, because they weren't all the same, and then fuzzy rule because I still don't know what is going on.

7 Neil …So what might be the rule here?

8 Danielle Rolling a die. [Neil: is that a rule?] Yes, but you still can't reproduce it. [Neil: So, what type of rule is that?] Clear (unsure)...

9 Neil Earlier you said that clear was a formula that we could follow but for fuzzy rules there was a level of what you called variability.

10 Danielle I'm really glad that you write what I say down, because I wouldn't have remembered (both laugh). So, yeah, this would still be fuzzy because there is still this level of variability. [Neil: We can roll the die, right?] Yeah, we can't guarantee what it's going to land on, so it would still be fuzzy. [Neil: Okay. So, the process would be, in terms of category] Oh, uh, stochastic.

As Danielle worked with Process Six, is was quickly able to determine that the process was not reproducible but replicable (line 6 of Excerpt 67). Danielle did have some hesitation over whether the rule to Process Six was clear or fuzzy (line 8). I believe that her hesitation was due to her image of rolling a die be a straightforward rule; she would know what to do. Danielle appeared to have lost track of the relationship between a clear rule and getting exact outcomes. However, after I reminded her of what she had said earlier, Danielle was able to identify the fuzzy nature of rolling a die and classify Process Six as stochastic (lines 9-10).

I then had Danielle work with Process Seven; the process that has both deterministic and stochastic aspects. Once again, Danielle started hunting for a pattern as I carried out the process (lines 1-4 of Excerpt 68). However, Danielle made a surprising move: she began to write down the values in an effort to stop thinking about patterns (line 6). Danielle appeared to realize that by focusing on whether the process was reproducible or replicable was both easier and more productive than trying to find a pattern from the start.

Excerpt 68. Danielle explores process seven

1	Neil	Process Seven. First result (4) [Danielle: Okay] Second result (5) [Danielle: okay] (Neil clicks, 2) [Danielle: Alright (laughs)] (Neil clicks, 3) [Danielle: okay] (Neil clicks, 1) So what do you think is going on?
2	Danielle	Well, I thought I knew, but now I'm not sure.
3	Neil	So what did you think.
4	Danielle	So I thought it would be like 4, 5, 2, 3, 0, 1, but now there's a one there. So, I'm not too sure. [Neil: so, you're thinking about patterns again.] Yeah, it always come down patterns. Okay. So [Neil: Ready?] Yeah. (Neil clicks, 2) Okay. (Neil clicks, 3) Alright. (Neil clicks, 1) Hold on, let me write all these down.
5	Neil	So why do you want to write them down?
6	Danielle	Cause if I stop thinking about patterns, all I need to know is if it is reproducible or replicable. Which means I would technically need to know a couple of them to test that out. [Neil: okay] (Danielle records several values on her sheet.) Okay. Can we fill it and restart it?...

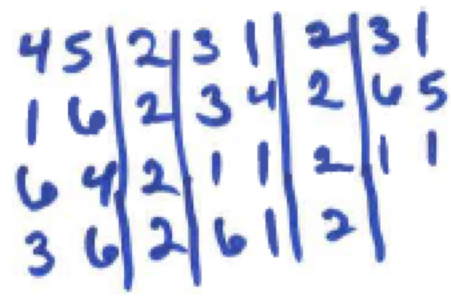


Figure 39. Danielle's four records of process seven

Danielle did not just write down the values for two times through Process Seven (Figure 39), she tracked four separate times through the process; three to convince herself to trust her intuitions (lines 1-7 of Excerpt 69) and a fourth at my request (lines 9-10).

Excerpt 69. Danielle explores process seven's repetition

1	Neil	Okay. So, I'll fill in. [Danielle: okay] Reset. Run the process the first time (1). [Danielle: okay] Second time (6). Third time (2). Fourth time (3). (Neil clicks, 4) (Neil clicks, 2) (Danielle records values on her paper) (Neil continues clicking, 6, 5) So what are you noticing?
2	Danielle	It is still one through six, but they are not the same. They are not identical. [Neil: they are not identical] mmm-Hmm.
3	Neil	So if we fill in...shall we do this again? [Danielle: Yes.] You seem a hesitant, like I'm (Danielle laughs)
4	Danielle	I can never be sure. [Neil: Can never be sure about what?] Whether I should just trust my gut instinct on this, or if I should just run this another time to be sure. [Neil: Well, what is it going to hurt run another time?] That's true.
5	Neil	So let's run this another time. [Danielle: Okay.] Are you ready? [Danielle: yeah] (Neil clicks, 6)
6	Danielle	Okay. (Neil clicks, 4) Yeah, so it is definitely different. [Neil: Are you sure?] yes. (Neil clicks, 2). (Neil clicks several times; 1, 1, 2) Yes, I'm sticking with my guns on this on. (laughs)
7	Neil	So write, go ahead and write down those numbers again. [Danielle: alright] I'll go one more that way...[Danielle: okay] Alright, we could fill in the rest right? [Danielle: mm-hmm (agreement)] We could run this again, right? Are you ready? [Danielle: yes] You seem very reluctant. [Danielle: yes] Why are you reluctant?
8	Danielle	Cause there's got to be like this nefarious plan...
9	Neil	No nefarious plan, I just want to make sure that, that you are very confident with (Danielle laughs) what you're claiming. [Danielle: okay] (Neil clicks, 3)

10 Danielle — Okay. (Neil clicks, 6, 2) (Sighs) So there is a pattern. [Neil: What do you mean that there's a pattern?] Sort of...so, not the next two, but the third from now should be a two. (Neil clicks three times, 6, 1, 2) Okay. (Neil fills the grid) So every third is a two.

In line 10 of Excerpt 69 Danielle notices that every third outcome of Process Seven is a two. This causes some issue for Danielle as she tried to fill in the worksheet (Figure 37) for Process Seven. I believe that my response to Danielle's question of the existence of middle options (line 2 of Excerpt 70) only clouded the issue at hand. While I had tried to clarify, I believe that had I given her the option to generate her own middle categories, I would have gotten better insight into how she was thinking about process seven. While Danielle was happy to pick both fixed and unfixed for outcomes (lines 9-11), she did not attempt to pick both for the type of repetition (lines 1-8) nor for the type of rule (line 12). Danielle saw both reproducible elements (line 4) and replicable elements (line 6). However, Danielle deferred to replicability as she noted that the pattern of a two every third outcome was replicable (line 6). Danielle's thinking here is in line with thinking that something that is reproducible is inherently replicable: replicability is about repeating the process, reproducibility is repeating the process and getting the same results. Danielle made use of the twos to establish the existence of fixed outcomes and a clear rule. I suspect that the regularity of the twos had Danielle thinking along the same lines that von Mises encouraged for looking for a gambling system. If an individual can find a gambling system, then he/she doesn't need to worry about anything else to make a judgement about the process. For making her final classification of Process Seven, Danielle went with a majority rules approach (lines 13-16). I do wonder what she would have come up with had I not forced her to choose between deterministic and stochastic, but rather if I let her create a new category.

Excerpt 70. Danielle classifies process seven

1	Neil	So are we replicable or reproducible? Or something in between?
2	Danielle	Closer to something in between. So, every third is guaranteed to be a two, but the ones in between those are variable. [Neil: okay] Do we have a middle option? [Neil: no.] Hmmm [Neil: At least not in terms of these categories (deterministic/stochastic)] Okay. [Neil: here (the three columns) we can create some middle categories] Okay.
3	Neil	But in terms of our categories here, we're either deterministic or we're stochastic. [Danielle: Okay] But right now, let's wrestle with these. [Danielle: okay] So in type of repetition, are we reproducible or replicable? Or both?
4	Danielle	Well, we can reproduce that every third number is going to be two.
5	Neil	Does that hold for the rest? [Danielle: For all of them?] For the whole sequence?
6	Danielle	Yes. You're making me doubt my sanity though. (Both chuckle) Yeah, it looks like it [Neil: okay] But the numbers in between those you can't, you can replicate the process, so I guess you can't [Neil: but we can only reproduce on the third] Yeah. [Neil: okay] So, like you can replicate the pattern on every third is going to be a two. [Neil: okay] Does that count?
7	Neil	So what are you going to fill in the box?
8	Danielle	I would it call it replicable. [Neil: Okay]
9	Neil	So fixed or unfixed outcomes?
10	Danielle	The twos are fixed. But everything else isn't. [Neil: But do we have at least have something that is fixed?] mmm-hmmm (agreement) Would every third number being fixed count as being a fixed outcome? [Neil: You tell me. It's your turn to make a decision.] (laughs) I don't like decision making. I'm not qualified for this. [Neil: yes, you are.] (laughs) Umm,
11	Neil	You can pick both. [Danielle: I'm allowed to pick both?] You're allowed to pick both in these (the three columns), you're not allowed to pick both in this (deterministic/stochastic column).
12	Danielle	Gotcha. I would say clear rule. [Neil: clear rule] yeah. Unfixed...
13	Neil	So in terms of category, what are we?
14	Danielle	What were the definitions between the two of them?
15	Neil	So, stochastic is unfixed, replicable, and fuzzy. All three have to be true. [Danielle: okay] Deterministic typically clear rules, typically reproducible, and fixed outcomes.

16 Danielle Alright, I'm going to go with two out of three, even technically either one is,...let's do deterministic. [Neil: okay]

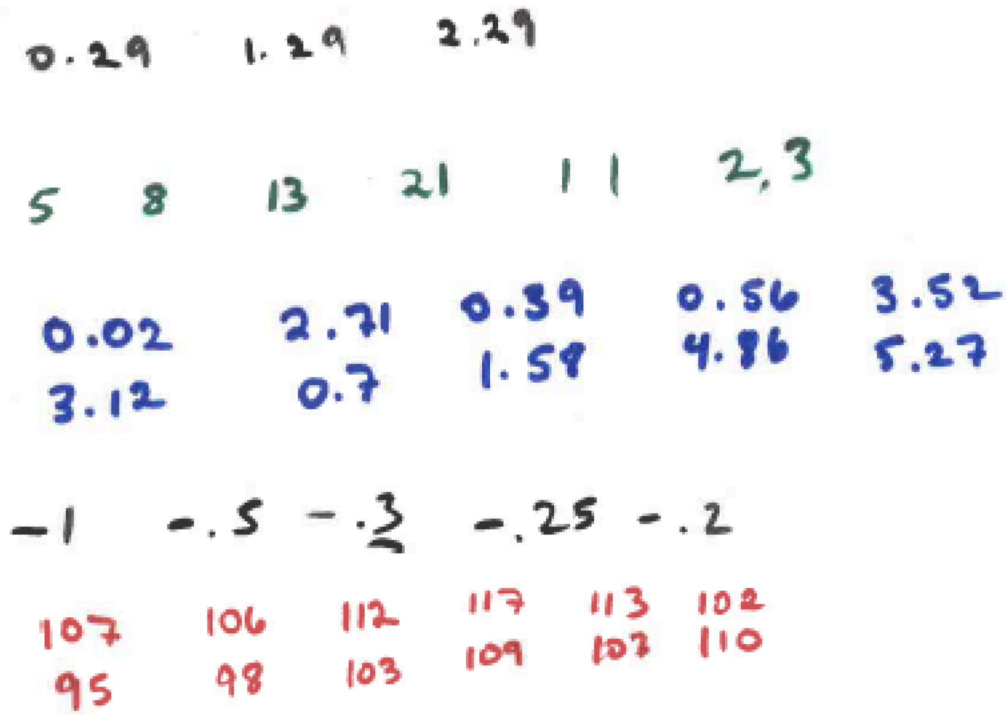


Figure 40. Danielle's records for the remaining processes

For all but one of the remaining six processes, Danielle's approach was the same: write several terms from the first time through the process. Then she would have me reset the applet and carry out the process again. Each color in Figure 40 denotes a different process, starting with Process Eight and skipping only Process Nine (deterministic, always –6). With the exception of Process Twelve (negative reciprocals of the natural numbers), Danielle had me carry out each process three times (Process Twelve was only twice). Her focus became checking for reproducibility or replicability first. Once she had established the type of repetition, she would know whether there were fixed or unfixed outcomes. She could then work on finding a pattern or clear rule for only those processes that were reproducible and had fixed outcomes. Danielle had developed a litmus test for whether she should devote time to looking for a pattern. Danielle viewed deterministic processes as processes that have a reproducible pattern/clear rule. For her, a stochastic processes had the hallmark of replicability and unfixed outcomes. Her litmus test was checking for replicability, a substantial shift away from her original method of looking for a pattern. At the start of the second session, I believe that Danielle's meaning for stochastic process as tied more closely to the absence of a pattern than to unfixed outcomes and replicability. However, by the end of the session, I believe that Danielle's meaning had shifted to prioritize the notion of replicability of a stochastic process over that of a pattern. Given the usefulness of her using her litmus test and having success, the notion of replicability appears to be a key component in developing a productive way of thinking about stochastic processes.

Chapter 9: Discussion and Conclusion

In this final chapter, I will return to my research question in light of what I have learned about Bonnie's and Danielle's thinking about stochastic processes. Additionally, I will discuss several limitations related to this research before proposing implications and future directions for research.

Comparing Bonnie and Danielle

The primary research question for this study is "What meanings for stochastic process do students develop during an instructional sequence based upon a hypothetical learning progression for thinking of distribution as describing the complete behavior of a stochastic process?" In answering this question, I found some distinct differences in the meanings that I believe Bonnie and Danielle developed.

For Bonnie, a stochastic process was any process that lacked any type of rule or pattern, being impossible to correctly guess what would happen next. Her conveyed meaning for randomness as being without pattern was the driver behind her meaning for stochastic process. When presented with a black-boxed process such was what she encountered with the Sequences applet, I believe that she thinks about the appearing sequence of outcomes and that the process is nothing more than the unveiling of those outcomes. Her thoughts focus on the sequence and looking for a set rule or pattern that would allow her to state what outcome comes next. When she cannot think up a set rule, then the sequence is random and so is the process of unveiling. Bonnie's way of thinking centers on the outcomes, not on the process itself and is evidenced by her pattern hunting. She is limited by her past experiences in looking for patterns. Bonnie mainly focused on arithmetic patterns and had trouble coordinating multiple patterns (e.g., Process 10—the shuffled Fibonacci Sequence).

Danielle's conveyed meaning shifted from centering on the lack of a pattern in the outcomes to looking at repeating the process the seeing if the outcomes are the same (values and order). If I were to present Danielle with a new black-boxed process, I believe that her thinking would entail having me unveil several outcomes, which she would record, and then have be reset and unveil more outcomes. She would then compare this second set to the prior set: if she found that the outcomes where the same values, in the same positions/order (i.e., reproducibility), she would declare that the underlying process is deterministic; otherwise she would say the process is replicable and stochastic. In Danielle's thinking there is a process that is more than unveiling; there is something behind the unveiling that can be repeated.

The thinking that Bonnie and Danielle employed led them to disagreements on only a few fronts; primary Processes One, Two, Seven, and Ten as shown in Table 12. For Process One, while both Bonnie's and Danielle's thinking revolved around pattern hunting, Bonnie was limited by her past experiences with primes. With support in a follow up portion at the end of the second exploratory teaching interview, Bonnie was able to make use of recognizing primes to revise her initial classification of Process One. For Process Two, Danielle came up with two different classifications based upon her imagining two different situations. I will discuss this more in the limitations section of the present chapter. The disagreement on Process Seven is a result of my forcing Danielle to choose between the two classifications. The disagreement on Process Ten stems from their thinking. Bonnie's thinking about classifying a process as deterministic or stochastic came down to her identifying a pattern or set rule. By the time Danielle saw this same process, her thinking revolved around whether the process was replicable or reproducible.

Table 12. Bonnie and Danielle's Classifications of the Sequences Applet Processes

Process	Bonnie	Danielle	Intended
Zero	Set Rule	Deterministic	Deterministic
One	Random	Deterministic	Deterministic
Two	Random	Deterministic, Stochastic	Stochastic
Three	Set Rule	Deterministic	Deterministic
Four	Set Rule	Deterministic	Deterministic
Five	Random	Stochastic	Stochastic
Six	Random	Stochastic	Stochastic
Seven	Random & Set Rule	Deterministic	Non-stochastic
Eight	Set Rule	Deterministic	Deterministic
Nine	Set Rule	Deterministic	Deterministic
Ten	Random	Deterministic	Deterministic
Eleven	Random	Stochastic	Stochastic
Twelve	Set Rule	Deterministic	Deterministic
Thirteen	Random	Stochastic	Stochastic

Bonnie and Danielle provide two cases for what meanings students can construct for stochastic process. Both of them had a traditional introductory statistics class, focusing on procedures. The curriculum at that time did not entail discussion of stochastic processes at all. Bonnie demonstrates that without explicit supports to develop a productive meaning for stochastic process, the role of stochastic process is nearly, if not entirely, absent from a student's thinking. The focus is entirely on the outcomes and the sequence that the student happened to observe. Danielle is a case where with support, a student could move away from only looking at a single set of trials of a process to comparing several sets of trials. The thinking that Danielle has the seeds for thinking about stochastic processes in a way that would support her in thinking about variation between collections (what others refer to as sample-to-sample variability): carrying out say a dozen trials of the process creates a sample, the resetting and getting a new dozen trials creates a second sample. Noticing the replicability of the process can then lead to questions of "how different are the samples?" and "why might the samples be different?". Both questions, and particularly the second, have the potential to help a student such as Danielle to begin thinking about stochastic processes less tied to particular outcomes.

A Third Participant's Thinking

Within both the description of Bonnie's and Danielle's work with the applets and their thinking, there is a third participant whose thinking evolved: me. At the onset of this work, my thinking about a stochastic [random] process was much more closely tied with my then meaning for randomness as sequence complexity. At that point in time, I believe that the process was a necessary nuisance to create a sequence which I would then check for the appearance of randomness by checking for term-based patterns, gambling systems, and complexity. However, through carrying out this research, my continued teaching of the concepts, and thinking through critiques, my thinking has evolved. At present, my thinking for a stochastic process centers much more on how I imagine the process unfolding and randomness is more the necessary nuisance.

In Chapter 4, I proposed an initial concept map for stochastic/random process (see Table 2) that held two top levels: first- and second-order stochastic processes. The essential difference was that a first-order stochastic process deals with the generation of raw data while the second-order stochastic process dealt with the process generating raw but also the process of using a statistic to learn something about the data collection.

While I stand by this distinction, I now believe that there is further refinement needed around the first-order stochastic process. At present, the first-order process conflates two elements in certain cases: the generative process of a data value and the process of selecting a member of a population. To illustrate this difference, consider the generative stochastic process of rolling any die that isn't designed to return a fixed value. The underlying process is the act of rolling the die and we can imagine that the fuzzy rule/inputs consist of the many factors we can imagine such as friction, force of the throw, the surface we're rolling the die on, manufacturing imperfections, etc. The die stops rolling, the end of this trial of the process, and we record the outcome. Alternatively, we could imagine the generative process of a person's height: the process consists of various biological and environmental factors making up the fuzzy rule/inputs. At a particular moment in time, we "freeze" the generative process and record the person's height. However, in addition to thinking about the generative process behind the person's height, we can also think about the selection process that led us to this person. We might make a sampling frame and carry out a simple random sampling procedure to get this particular person. Teasing apart the generative and selective aspects of a stochastic process should provide new opportunities to continue investigating students' understandings of stochastic processes. One place that such a teasing apart could bear fruit is in the notion of fixed versus random effects in regression models. In my current thinking a "fixed effect" would be a case where I suppress any thought about the generative process behind the term and only go with the selective process. However, if I were to think about the factor as a "random effect", I am now starting to bring both the selective *and* generative processes together.

Implications-Concept Map

Both Bonnie's and Danielle's meanings for stochastic process are missing from the original concept map for stochastic/random processes (Table 2). Additionally, my recent thinking about generative and selective processes involved at the first-order stochastic process level is also missing. Together, these have led me to develop a revised concept map as shown in Table 13.

The first addition is the level of "Model Without a Pattern". This way of thinking about a stochastic process is what underpins Bonnie's meaning. In such a case, the individual views a stochastic process as any process for which a pattern cannot be found. While the individual will carry out several trials of the process, the individual does not think to start all over and re-carry out the process. I suspect that such a meaning for stochastic process is irrevocably tied with the lack of a discernable pattern meaning for randomness (see Table 10).

The second addition is the "Replicable Model" level to the concept map. This particular way of thinking goes beyond chance models in that the student does not bring up statistical fairness. However, this way of thinking does not quite reach the level of first-order stochastic model in that the individual does not engage in checking for the appearance of randomness. Rather the individual focuses on whether the process can be repeated and if the outcomes are exactly the same (i.e., the process is reproducible) or if

Table 13. Revised Stochastic/Random Process Concept Map

<table>
<tr><td colspan="2">Second-Order Random Models
The student imagines a method of generating values of a statistic of interest to answer a research question about a population. In generating the values of the statistics, the student envisions a first-order random model to get data values necessary for the generation of the value the statistic. The student anticipates being able to repeat this larger process of generating values of the statistic indefinitely and that the values will not always be the same. Adapted from (Liu & Thompson, 2002).</td></tr>
<tr><td colspan="2">First-Order Random Model
The student imagines a method of generating values of a random variable of interest to answer a question about that random variable in some population. The student's image includes the carrying out the method infinitely many times and expecting variation in outcomes of each trial of the method. The outcomes are values of the random variable and form the sequence that the student checks for randomness. Adapted from (von Mises, 1981).</td></tr>
<tr><td>Generative Process
The student imagines a loosely defined mechanism that would result a datum at the conclusion of carrying out the process once. The student's model might include various factors but might not proposes detailed operations on those factors. For example, the student might discuss how parents' height or age of the child impact the child's height but not be able to give an explicit model.</td><td>Selective Process
The student imagines a way of picking members out of a population. The end result of the process is the selected object/living being and the values of whichever attributes he/she might be interested in.</td></tr>
<tr><td colspan="2">Replicable Model
The student imagines a process that he/she can repeat for a set number of trials as well as anticipating that he/she can then carry out new iterations of these sets of trials. The student's image of process is such that the process will show replicability (repeating the process and getting like outcomes) but not reproducibility (repeating the process and getting the exact same outcomes). This way of thinking allows the student to also identify whether the process will produce fixed or unfixed outcomes and have a clear or fuzzy rule.</td></tr>
<tr><td colspan="2">Chance Model
The student imagines a method of generating values of the random variable, while the student could carry out an infinite number of times, the student does not feel the need to carry out any trials of the method in order to answer questions about the random variable. The student is able to completely specify each and every outcome without running trials and the student assumes that each outcome has the same chance of occurring as every other outcome. Adapted from (von Mises, 1981; Weisberg, 2014).</td></tr>
<tr><td colspan="2">Model Without a Pattern
The student imagines a process that does not follow any particular rule or pattern. The student engages in pattern hunting to evaluate whether or not the process follows, but the student does not think to carry out the process with new iterations of different trials.</td></tr>
<tr><td colspan="2">Deterministic Model
The student imagines a process where he/she anticipates what the result will be before carrying out the process. Further, the student anticipates that if he/she carries out the process under identical conditions again and again, the result will be essentially the same each time.</td></tr>
<tr><td colspan="2">Null Model
The student believes that there is no way to model the behavior of a random variable or statistic.</td></tr>
</table>

the outcomes are different (i.e., the process is replicable). Given Danielle's success, this particular meaning for stochastic process is both useful in the moment and productive. Danielle was able to use this meaning to have success in identifying which processes where deterministic or stochastic. This meaning allowed her to bypass hunting for patterns by recognizing when such an activity would not be fruitful (i.e., finding replicable processes). From an instructional point of view, this meaning is productive in that once a student has this meaning for a stochastic process, helping the student develop methods for checking the appearance of randomness that become integrated with this meaning will move the student up a level in the construct map.

Given her responses, I believe that Bonnie is at the Model Without a Pattern level. Bonnie was able to operate successfully in her introductory statistics course with such a meaning and I suspect that if she were to take a second course, she would continue to have success with this meaning. With her imagining that stochastic process is a process that does not have a set rule or pattern to the outcomes, her understanding of distribution will be limited. I hypothesize that her meaning for distribution will be along the lines of the D.S. Moore definition (i.e., a distribution is the domain and how often each value occurs). While Moore might intend for students to understand "how often each value occurs" to refer to probability, I suspect students like Bonnie will interpret that phrase as being about absolute frequency. Further, I suspect that given the primacy that the lack of a pattern has in Bonnie's thinking, her image of distribution will not support her in using distributions to make claims about long-run behavior.

Danielle's view of stochastic process as centering on the notion of replicability provides her with an escape from the tyranny of patterns. While her thinking allowed her to check off the other two aspects (unfixed outcomes and fuzzy rule), I believe that the real potential of this way of thinking was untapped in this study. I suspect that Danielle's meaning for stochastic process offers the potential for spring boarding into investigating whether there is any kind of regularly to the process. While she only looked at a few trials each time she checked for replicability, pushing her to look much larger sequences, with the assistance of data visualizations could help her start to see the emergent pattern of long-run behavior, that is to say, the distribution.

One activity that could help Danielle here is to consider the selection of US adult and measure their heights in inches. Figure 41 shows three moments: left-after carrying out the process eleven times; middle-after carrying out the process 111 times; right-after carrying out the process 10,111 times. After the right panel (or even examining the graph while process continues to run), a conversation about what Danielle might notice about the graph could be had. Then, Danielle could be asked what she think would happen if

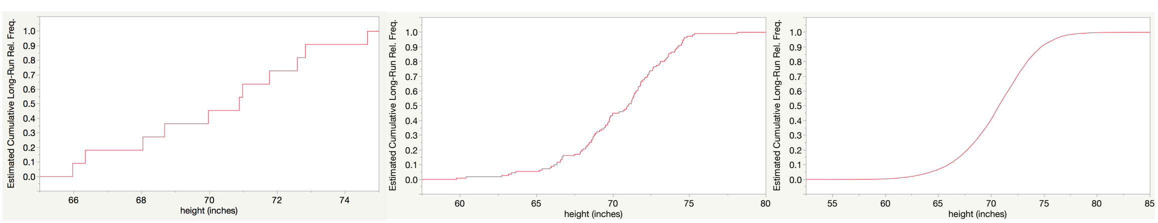


Figure 41. Three moments from a simulation

we cleared out all of the data and started all over: what would she anticipate seeing. Figure 42 shows two additional panels of carrying out the simulation another 10,111 times. I suspect that Danielle would be able to argue that the sequences of data are not identical, but that there was some commonality. Asking her what might be driving the commonality could provide an opening to not only discuss distribution but also get Danielle to start thinking about a generative stochastic process.

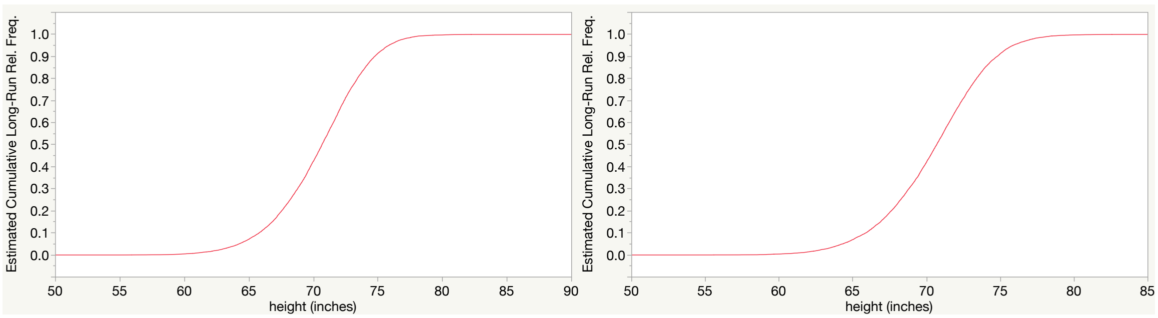


Figure 42. Two additional graphs from a simulation

Limitations

An important set of limitations revolves around the tasks that I used within the research. My meanings for the ideas of randomness and stochastic process evolved over the course of the study. However, the meanings that I had held at the start dictated how I originally conceived and then made the Four Rooms and Sequences applets and their tasks. In particular, my meaning for randomness centered on examining a sequence of outcomes and checking whether there was term-based pattern, a gambling system and/or sufficient complexity. Further, my meaning for stochastic process was much more tied up with my meaning for randomness. Looking back, I would describe my meaning for stochastic process as anything process that would generate a sequence where I could check the boxes for no term-based pattern, no gambling system, and sufficiently complex descriptions. The roles of unfixed outcomes, fuzzy rules, and replicability minor aspects pushed to the far edges of the stage by my meaning for randomness. These meanings directly impacted the choices I made in the design of the underlying processes that I used for the rooms and the entire design of the Sequences Applet.

I originally designed the Four Rooms applet to serve as an introduction to the role of randomness and stochastic processes within the familiar context of using a light switch. I intended for the prompt asking the student to describe how they would create a replacement button/switch for each of the four rooms as the opportunity for them to start making distinctions between different kinds of processes. However, the flashing lights of the four rooms and looking for patterns in those lights appears to support students in focusing on the outcomes of the generative processes rather than the processes themselves.

The Sequences applet, as the very name implies, is a consequence of my sequence complexity meaning for randomness. While I had planned for explicit details of the generative process to be absent, my design choices left the generative processes out in the cold. Thus, I am limited in teasing apart Bonnie and Danielle's thinking about the pattern of outcomes and what they actually conceived as the generative process. There is one exception to this: Danielle's work with Process Two of the Sequences applet (see Excerpt 63, Excerpt 64, and Excerpt 65). When Danielle recognized the outcomes as being nucleotides, she assimilated the partial sequence to her scheme for DNA sequencing. I supported her assimilation and even prompted her to conceive of sequencing and reading out a single person's DNA. She then used her understanding of this generative process to make decisions about fixed vs. unfixed outcomes, reproducible vs. replicable, and clear or fuzzy rules. When I asked Danielle how we could check whether the sequence was reproducible, she went back to the generative process of sequencing the person's DNA. When I re-ran Process Two in the Sequences applet, the values were not the same as what she originally saw. Danielle game up with two tweaks to her original generative process: that there was an error made in sequencing the DNA or that the DNA had changed. Both of these options provided an unseized opportunity to dig into Danielle's image of stochastic process within an outcome. As I reflect upon this episode, I am curious as to how Danielle comment on a proposed process of selecting different individuals and recording the first nucleotide on the same chromosome. With the remaining processes in the Sequence's applet, I cannot separate either student's thinking about the underlying generative process from the displayed sequence of outcomes.

Future Directions

There are several lessons that I and others can take away from this study. First, the set of instructional activities and interventions do appear to support students in developing productive meanings for stochastic process. The Four Rooms activity appears to support students in shifting away from the lower end of the randomness construct map (Table 10) and towards a meaning of the lack of a discernable pattern. Additional work on developing activities to take students from the lack of a pattern meaning to the notion that randomness is an attribute of process is needed. The Sequences Applet with the intervention that Danielle received supports students in developing the replicability image for stochastic process. Marrying the sequences applet with instruction on checking for the appearance of randomness through term-based patterns, Kolmogorov's Sequence Complexity, and von Mises's Principle of the Impossibility of a Gambling System could provide students with experiences to help them develop a meaning consistent with a first-order stochastic process.

Second, this study highlights that Statistics Education needs to take seriously the meanings that students bring with them for core concepts such as randomness and stochastic process. All too often introductory statistics textbooks do not attend to students' meanings for these concepts and in some cases, do not support students in developing any meaning. For example, Agresti, Franklin, and Klingenberg (2017) state "Random is often thought to mean chaotic or haphazard, but randomness is an extremely powerful tool for obtaining good samples and conducting experiments" (pg. 12). Later they present randomness as "randomly assigning subjects to treatments or randomly selecting people for a sample" (Agresti et al., 2017, p. 201). In their discussion of random sampling, Agresti et al. allow for students to continue using the notions of haphazard or unplanned that Kaplan et al. (2009) as well as the notion of sampling by chance from Kaplan et al. (2014). While Agresti et al. (2017) do discuss some stochastic processes in their until on probability (e.g., rolling dice), their emphasis is on the outcomes, not on the underlying process. Figure 43 shows how these authors connect probability, randomness, and random phenomena (meaning outcomes); their next section heading is "Long-Run Behavior of Random Outcomes" (Agresti et al., 2017, p. 203).

I proposed two secondary research questions as part of this study:

- What impact do students' meanings for randomness, random variable, and probability have on the development of their meaning for stochastic process during the instructional sequence?
- What images of accumulation do students develop during the instructional sequence?

I did not investigate the students' images of accumulation with any rigor. At best, the students' notions of accumulation are embedded in the various applets. For example, the Sequences applet takes care of the accumulation by leaving a record, thus releasing the students from having to do that work. While I did collect students' responses about random variables and probability during the clinical interview portion (see the Appendix), I did not analyze that data for this dissertation. However, I can answer the question of the impact of students' meanings for randomness. Absent the interventions Danielle received to get her to focus on replicability, unfixed outcomes, and fuzzy rules, a student's meaning for randomness appears to be a major contributor to his/her meaning for a stochastic process. Bonnie is an excellent example of this phenomenon.

While there are some positive aspects (e.g., discussing probability as a long-run proportion), these authors do not attempt to problematize problematic meanings for randomness that students bring with them. Nor do they successfully help students develop an image of stochastic processes; rather the support students on being focused on the outcomes. I believe that the approach used by Agresti et al. supports students in engaging in pattern hunting behavior and at best, developing a meaning consistent with Bonnie's for stochastic processes (the lack of a pattern).

Lock, Lock, Lock Morgan, Lock and Lock (2017) present the concept of randomness to students in two ways: as a way to prevent sampling bias and as the idea of uncertainty. They provide students with a clear-cut caution (Figure 44) that students should not think of randomness as haphazard. While this is improvement on Agresti et al. (2017), Lock et al. (2017) do not go on to discuss how students should think about randomness outside of random sampling within the main part of their book. While Lock et al. make use a simulation approach to teaching statistical inference, they do not discuss stochastic processes until Chapter P; their probability rules chapter located just before Appendix A. "A process is *random* if its outcome is uncertain" (Lock et al., 2017, p. 690, emphasis in original). Again, I suspect that students using the Lock et al. book will develop a meaning for stochastic process consistent with the lack of a pattern.

In everyday life, often you must make decisions when you are uncertain about the outcome. Should you invest money in the stock market? Should you get extra collision insurance on your car? Should you start a new business, such as opening a pizza place across from campus? Daily, you face even mundane decisions, such as whether to carry an umbrella with you in case it rains.

This chapter introduces **probability**—the way we quantify uncertainty. You'll learn how to measure the chances of the possible outcomes for **random phenomena**—everyday situations for which the outcome is uncertain. Using probability, for instance, you can find the chance of winning the lottery. You can find the likelihood that an employer's drug test correctly detects whether you've used drugs. You can measure the uncertainty that comes with randomized experiments and with random sampling in surveys. The ideas in this chapter set the foundation for how we'll use probability in the rest of the book to make inferences based on data.

5.1 How Probability Quantifies Randomness

As we discovered in Chapter 4, statisticians rely on an essential component to avoid bias in gathering data. This is **randomness**—randomly assigning subjects to treatments or randomly selecting people for a sample. Randomness also applies to the outcomes of a response variable. The possible outcomes are known, but it's uncertain which outcome will occur for any given observation.

We've all employed randomization in games. Some popular randomizers are rolling dice, spinning a wheel, and flipping a coin. Randomization helps to make a game fair, each player having the same chances for the possible outcomes. Rolls of dice and flips of coins are simple ways to represent the randomness of randomized experiments and sample surveys. For instance, the head and tail outcomes of a coin flip can represent drug and placebo when a medical study randomly assigns a subject to receive one of two treatments.

In Words

Phenomena are any observable occurrences.

With a *small* number of observations, outcomes of *random phenomena* may look quite different from what you expect. For instance, you may expect to see a random pattern with different outcomes; instead, exactly the same outcome may happen a few times in a row. That's not necessarily a surprise; unpredictability for any given observation is the essence of randomness. We'll discover, however, that with a *large* number of observations, summary statistics settle down and get increasingly closer to particular numbers. For example, with 4 tosses of a coin, we wouldn't be surprised to find all 4 tosses resulting in heads. However, with 100 tosses, we would be surprised to see all 100 tosses resulting in heads. As we make more observations, the proportion of times that a particular outcome occurs gets closer and closer to a certain number we would expect. This long-run proportion provides the basis for the definition of *probability*.

Figure 43. Agresti et al. (2017, pg. 201) relating probability, randomness, and random phenomena

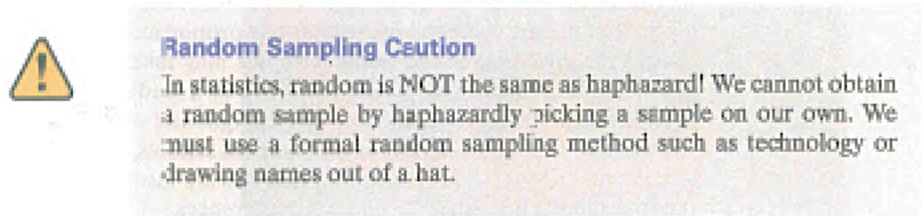

Random Sampling Caution
In statistics, random is NOT the same as haphazard! We cannot obtain a random sample by haphazardly picking a sample on our own. We must use a formal random sampling method such as technology or drawing names out of a hat.

Figure 44. A caution on how not to think about randomness (Lock, Lock, Lock Morgan, Lock, & Lock, 2017, p. 21)

However, there is some hope for the activities in this study fitting in with at least one curriculum on the market. The text *Introduction to Statistical Investigations* (Tintle et al., 2016) places an early emphasis on random processes. In discussing the game show *Let's Make a Deal*, these authors write "this game is an example of a ***random process***: Although the outcome for an individual game is not known in advance, we expect to see a very predictable pattern in the results if you play this game many, many times" (Tintle et al., 2016, p. 10, emphasis in original). After having the students play *Let's Make a Deal* a number of times, the authors provide the following recap:

> A ***random process*** is one that can be repeated a very large number of times (in principle, forever) under identical conditions with the following property: Outcomes for any one instance cannot be known in advance, but the proportion of times that particular outcomes occur in the long run can be predicted. (Tintle et al., 2016, p. 13)

This text appears to lay the groundwork for helping students develop a first-order random model meaning for stochastic process. While Tintle et al. lay this foundation in their preliminaries chapter, in chapter one they shift to discussing chance models, consistent with how I have used this phrase in Table 13 (i.e., assuming statistical fairness).

Conceptualizing distribution as the long-run behavior of a process provides students with a powerful framework to reason probabilistically and to engage in statistical inference. However, central to this conceptualization is the notion of stochastic process. Students must be able to imagine that stochastic/random process is one that can be repeated infinitely, be replicable but not reproducible, have unfixed outcomes, have a fuzzy rule, and the attribute of randomness (i.e., the minimization of bias and allowing for long-run predications).

REFERENCES

Agresti, A., Franklin, C. A., & Klingenberg, B. (2017). *Statistics: the art and science of learning from data* (Fourth edition). Boston: Pearson.

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., … Witmer, J. A. (2005). *Guidelines for assessment and instruction in statistics education: College report*. Alexandria, VA: American Statistical Association. Retrieved from http://it.stlawu.edu/~rlock/gaise/GAISECollege.pdf

Arnold, P., & Pfannkuch, M. (2012). The language of shape. In *Proceedings of the 12th International Congress on Mathematical Education (ICME-12, July, 2012), Seoul, South Korea* (pp. 2446–2455). Retrieved from http://www.aucklandmaths.org.nz/wp-content/uploads/2013/05/pip-arnolds-research.pdf

Arnold, P., & Pfannkuch, M. (2014). Describing distributions. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education*. Flagstaff, AZ: International Statistical Institute. Retrieved from http://icots.info/icots/9/proceedings/pdfs/ICOTS9_8G1_ARNOLD.pdf

Bakker, A., & Gravemeijer, K. P. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147–168). Springer. Retrieved from http://link.springer.com/content/pdf/10.1007/1-4020-2278-6_7.pdf

Batanero, C., Green, D. R., & Serrano, L. R. (1998). Randomness, its meanings and educational implications. *International Journal of Mathematical Education in Science and Technology*, *29*(1), 113–123. https://doi.org/10.1080/0020739980290111

Batanero, C., & Serrano, L. (1999). The Meaning of randomness for secondary school students. *Journal for Research in Mathematics Education*, *30*(5), 558. https://doi.org/10.2307/749774

Beichelt, F., & Fatti, L. P. (2002). *Stochastic processes and their applications*. Boca Raton, Fla: CRC Press.

Bennett, D. J. (1993). The development of the mathematical concept of randomness: Educational implications (dissertation). New York University, New York.

Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal*, *3*(2), 42–63.

Blumer, H. (1986). *Symbolic interactionism: perspective and method*. Berkeley: University of California Press.

Carver, R., Everson, M., Gabrosek, J., Horton, N. J., Lock, R., Mocko, M., … Wood, B. (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016*. Alexandria, VA: American Statistical Association.

Clark, J., Kraut, G., Mathews, D., & Wimbish, J. (2007). *The fundamental theorem of statistics: Classifying student understanding of basic statistical concepts*. Retrieved from http://www1.hollins.edu/faculty/clarkjm/stat2c.pdf

Cobb, G. (2015). Mere renovation is too little too late: We need to rethink our undergraduate curriculum from the ground up. *The American Statistician*, *69*(4), 266–282. https://doi.org/10.1080/00031305.2015.1093029

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, *104*(9), 801–823.

Cobb, P., Confrey, J., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9–13.

Cobb, P., Jackson, K., & Dunlap, C. (2016). Design research: An analysis and critique. In L. D. English & D. Kirshner (Eds.), *Handbook of International Research in Mathematics Education* (3rd ed., pp. 481–403). New York: Routledge.

Corcoran, T., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An Evidence-based approach to reform* (CPRE No. RR-63) (p. 86). New York, NY: Columbia University. Retrieved from http://eric.ed.gov/?id=ED506730

De Finetti, B. (1974). Theory of probability; a critical introductory treatment. London, New York: Wiley.

Derry, S. J. (1996). Cognitive schema theory in the constructivist debate. *Educational Psychologist*, *31*(3–4), 163–174. https://doi.org/10.1080/00461520.1996.9653264

Dewey, J. (1910). *How we think*. Boston: D. C. Heath & Co.

Diez, D. M., Barr, C. D., & Çetinkaya-Rundel, M. (2016). *OpenIntro Statistics* (3rd ed.). OpenIntro.

Doerr, H. M. (2000). How can I find a pattern in this random data? *The Journal of Mathematical Behavior*, *18*(4), 431–454. https://doi.org/10.1016/S0732-3123(00)00023-7

Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: a review and analysis. *Studies in Science Education*, *47*(2), 123–182. https://doi.org/10.1080/03057267.2011.604476

Dwass, M. (1962). A Fluctuation theorem for cyclic random variables. *The Annals of Mathematical Statistics*, *33*(4), 1450–1454.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Falk, R. (1991). Randomness-an ill-defined but much needed concept. *Journal of Behavioral Decision Making*, *4*(3), 215–218.

Falk, R., & Konold, C. (1994). Random means hard to digest. *Focus on Learning Problems in Mathematics*, *16*(1), 2–12.

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

Faradj, M. K. (2004). *Which mean do you mean? An Exposition on means*. Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.195.3700&rep=rep1&type=pdf

Fielding-Wells, J. (2014). Where's your evidence? Challenging young students' equiprobability bias through argumentation. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education*. Flagstaff, AZ: International Statistical Institute. Retrieved from http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_2B2_FIELDINGWELLS.pdf

Franklin, C. A., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: a pre-K--12 curriculum framework*. Alexandria, VA: American Statistical Association.

Garfield, J., & Ben-Zvi, D. (2004). Research on statistical literacy, reasoning, and thinking: Issues, challenges, and implications. In *The challenge of developing statistical literacy, reasoning and thinking* (pp. 397–409). Springer.

Goldin, G. A. (2000). A scientific perspective on structured, task-based interviews in mathematics education based research. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 517–545). Mahwah N.J.: Lawrence Erlbaum Associates, Inc.

Goldstein, D. G., & Rothschild, D. (2014). Lay understanding of probability distributions. *Judgment and Decision Making*, *9*(1), 1–14.

Green, D. (1987). Probability concepts: Putting research into practice. *Teaching Statistics*, *9*(1), 8–14. https://doi.org/10.1111/j.1467-9639.1987.tb00613.x

Hatfield, N. J. (2013). The Action, process, object, and schema Theory for sampling/sampling distribution. In S. Brown, G. Karakok, K. H. Roh, & M. Oehrtman (Eds.), *Proceedings of the 16th Annual Conference on Research in Undergraduate Mathematics Education* (Vol. 1, pp. 296–307). Denver, Colorado: Mathematical Association of America.

Hatfield, N. J. (2016a). An initial look at students' conveyed meanings for probability. In T. Fukawa-Connelly, N. Infante, M. Wawro, & S. Brown (Eds.), *Proceedings of the 19th Annual Conference on Research in Undergraduate Mathematics Education* (pp. 190–203). Pittsburgh, PA: Mathematical Association of America.

Hatfield, N. J. (2016b, February). *Probabilistic thinking: An initial look at students' meanings for probability*. Contributed research presented at the 19th Annual Conference on Research in Undergraduate Mathematics Education, Pittsburgh, PA.

Horvath, J. K., & Lehrer, R. (1998). A model-based perspective on the development of children's understanding of chance and uncertainty. In S. P. Lajoie (Ed.), *Reflections on Statistics: Learning, Teaching, and Assessment in Grades K-12* (pp. 121–148). Mahwah N.J.: Lawrence Erlbaum Associates.

Illowsky, B., Dean, S., & OpenStax College. (2013). *Introductory statistics*. OpenStax College. Retrieved from http://cnx.org/content/col11562/latest/

Jacob, B. L., & Doerr, H. M. (2014). Statistical reasoning with the sampling distribution. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education*. Flagstaff, AZ: International Statistical Institute.

Jonckheere, A. R., Mandelbrot, B. B., & Piaget, J. (1958). *La lecture de l'expérience [Observation and decoding of reality]*. Paris: Presses Universitaires De France.

Kahneman, D., & Tversky, A. (1974). Subjective probability: A judgment of representativeness. In C.-A. S. Staël von Holstein (Ed.), *The Concept of Probability in Psychological Experiments* (pp. 25–48). Boston, MA: D. Reidel Publishing Company. Retrieved from http://link.springer.com/chapter/10.1007/978-94-010-2288-0_3

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, *11*(2), 143–157. https://doi.org/10.1016/0010-0277(82)90023-3

Kaplan, J. J., Fisher, D. G., & Rogness, N. T. (2009). Lexical ambiguity in statistics: What do students know about the words association, average, confidence, random and spread. *Journal of Statistics Education*, *17*(3).

Kaplan, J. J., Rogness, N. T., & Fisher, D. G. (2014). Exploiting lexical ambiguity to help students understand the meaning of random. *Statistics Education Research Journal*, *13*(1), 9–24.

Kennedy, C. A., & Wilson, M. (2007). *Using progress variables to interpret student achievement and progress* (BEAR Technical Report No. 2006-12–01). University of California, Berkeley. Retrieved from http://drupal.bear.berkeley.edu/sites/default/files/Kennedy_Wilson2007.pdf

Kolmogorov, A. N. (2013). *Foundations of the theory of probability*. Martino Fine Books.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, *6*, 59–98. https://doi.org/10.1207/s1532690xci0601_3

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning*, *12*(3), 217–230. https://doi.org/10.1007/s10758-007-9123-1

Kuzmak, S. (2016). What's missing in teaching probability and statistics: Building cognitive schema for understanding random phenomena. *Statistics Education Research Journal*, 179.

Leavy, A. M. (2006). Using data comparison to support a focus on distribution: Examining preservice teacher's understandings of distribution when engaged in statistical inquiry. *Statistics Education Research Journal*, *5*(2), 89–114.

Lecoutre, M.-P., Durand, J.-L., & Cordier, J. (1990). A Study of two biases in probabilistic judgments: Representativeness and equiprobability. In *Advances in Psychology* (Vol. 68, pp. 563–575). Elsevier. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0166411508613436

Lehrer, R. (2013). A learning progression emerges in a trading zone of professional community and identity. In R. Mayes & L. L. Hatfield (Eds.), *Quantitative reasoning in mathematics and science education: Papers from an international STEM research symposium* (Vol. 3, pp. 173–186). Laramie, WY: University of Wyoming.

Lehrer, R., & Schauble, L. (2002). Distribution: A resource for understanding error and natural variation. *International Association for Statistical Educations*. Retrieved from https://www.stat.auckland.ac.nz/~iase/publications/1/8b3_lehr.pdf

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal*, *41*(3), 635–679.

Lipson, K. (2003). The role of the sampling distribution in understanding statistical inference. *Mathematics Education Research Journal*, *15*(3), 270–287.

Liu, Y., & Thompson, P. W. (2002). Randomness: Rethinking the foundation of probability. In D. Mewborn (Ed.). Presented at the Proceedings of the Twenty-fourth Annual Meeting of the International Group for the Psychology of Mathematics Education, PME-NA.

Liu, Y., & Thompson, P. W. (2007). Teachers' understandings of probability. *Cognition and Instruction*, *25*(2–3), 113–160.

Lock, R. H., Lock, P. F., Lock Morgan, K., Lock, E. F., & Lock, D. F. (2017). *Statistics: unlocking the power of data*. John Wiley & Sons, Inc.

Magalhães, M. N. (2014). Challenges for learning about distributions in courses for future mathematics teachers. In *Sustainability in statistics education*. Flagstaff, AZ: International Statistical Institute. Retrieved from http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_6F2_MAGALHAES.pdf

Mathews, D., & Clark, J. M. (2007). *Successful students' conceptions of mean, standard deviation, and the central limit theorem*. Unpublished. Retrieved from http://www1.hollins.edu/faculty/clarkjm/stats1.pdf

Mead, G. H. (1910). Social consciousness and the consciousness of meaning. *Psychological Bulletin*, *7*(12), 397.

Mead, G. H. (1912). The mechanism of social consciousness. *The Journal of Philosophy, Psychology and Scientific Methods*, *9*(15), 401–406.

Metz, K. E. (1998). Emergent understanding and attribution of randomness: Comparative analysis of the reasoning of primary grade children and undergraduates. *Cognition and Instruction*, *16*(3), 285–365. https://doi.org/10.2307/3233647

Montangero, J., & Maurice-Naville, D. (1997). *Piaget, or, The advance of knowledge*. Mahwah, N.J: Lawrence Erlbaum Associates, Inc.

Moore, D. S., McCabe, G. P., & Craig, B. A. (2012). *Introduction to the practice of statistics* (7th ed.). New York, NY: W. H. Freeman.

Moore, D. S., McCabe, G. P., & Craig, B. A. (2015). *Introduction to the practice of statistics* (Eighth edition/Student edition). New York: W.H. Freeman and Company, a Macmillan Higher Education Company.

Moore, D. S., McCabe, G. P., & Craig, B. A. (2017a). Looking at data--relationships [PowerPoint slides]. In *Introduction to the Practice of Statistics* (9th ed.). Macmillan Learning. Retrieved from https://macmillanlearning.com/Catalog/product/introductiontothepracticeofstatistics-ninthedition-moore/instructorresources

Moore, D. S., McCabe, G. P., & Craig, B. A. (2017b). Probability: The study of randomness [PowerPoint slides]. In *Introduction to the practice of statistics* (9th ed.). Macmillan Learning. Retrieved from https://macmillanlearning.com/Catalog/product/introductiontothepracticeofstatistics-ninthedition-moore/instructorresources

Moore, K. C., & Thompson, P. W. (2015). Shape thinking and students' graphing activity. In *Proceedings of the 18th Meeting of the MAA Special Interest Group on Research in Undergraduate Mathematics Education* (pp. 782–789). Pittsburgh, PA: Mathematical Association of America. Retrieved from http://pat-thompson.net/PDFversions/2015MooreShapeThinking.pdf

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics : an overview.* Reston, VA: National Council of Teachers of Mathematics.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, D.C.: National Governors Association Center for Best Practices, Council of Chief State School Officers. Retrieved from www.corestandards.org

Oxford English Dictionary. (2015, December). "distribution, n.". Retrieved January 21, 2016, from http://www.oed.com/view/Entry/55781?redirectedFrom=distribution

Peck, R., Gould, R., & Miller, S. J. (2013). *Developing essential understanding of statistics for teaching mathematics in grades 9-12*. (P. S. Wilson, Ed.). Reston, VA: National Council of Teachers of Mathematics.

Piaget, J. (1995). *The essential Piaget* (100th Anniversary ed). Northvale, N.J: J. Aronson.

Piaget, J., & Elkind, D. (1968). *Six psychological studies*. New York: Vintage Books.

Prodromou, T. (2012). Students' construction of meanings about the co-ordination of the two epistemological perspectives on distribution. *International Journal of Statistics and Probability*, *1*(2). https://doi.org/10.5539/ijsp.v1n2p283

Reading, C., & Canada, D. (2011). Teachers' knowledge of distribution. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (Vol. 14, pp. 223–234). Dordrecht: Springer Netherlands. Retrieved from http://link.springer.com/10.1007/978-94-007-1131-0_23

Ross, S. M. (2010). *Introduction to probability models* (10th ed). Amsterdam ; Boston: Academic Press.

Saldanha, L. A. (2016). Conceptual issues in quantifying unusualness and conceiving stochastic experiments: Insights from students' experiences in designing sampling simulations. *Statistics Education Research Journal*, 81.

Saldanha, L. A., & Liu, Y. (2014). Challenges of developing coherent probabilistic reasoning: Rethinking randomness and probability from a stochastic perspective. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting Plural Perspectives* (pp. 367–396). Netherlands: Springer.

Saldanha, L. A., & Thompson, P. W. (2003). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, *51*(3), 257–270.

Saldanha, L. A., & Thompson, P. W. (2007). Exploring connections between sampling distributions and statistical inference: An analysis of students' engagement and thinking in the context of instruction involving repeated sampling. *International Electronic Journal of Mathematics Education*, *2*(3), 270–297.

Saldanha, L. A., & Thompson, P. W. (2014). Conceptual issues in understanding the inner logic of statistical inference: Insights from two teaching experiments. *The Journal of Mathematical Behavior*, *35*, 1–30. https://doi.org/10.1016/j.jmathb.2014.03.001

Samuels, M. L. (2015). *Statistics for the life sciences* (5th ed.). Boston: Pearson Education.

Savage, L. J. (1972). *The foundations of statistics* (2d rev. ed). New York: Dover Publications.

Sheats, R. D., & Shane Pankratz, V. (2002). Understanding distributions and data types. *Seminars in Orthodontics*, *8*(2), 62–66. https://doi.org/10.1053/sodo.2002.32075

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, *26*(2), 114–145.

Steffe, L. P. (1983). Children's algorithms as schemes. *Educational Studies in Mathematics*, *14*(2), 109–125. https://doi.org/10.2307/3482541

Steffe, L. P. (1992). Schemes of action and operation involving composite units. *Learning and Individual Differences*, *4*(3), 259–309. https://doi.org/10.1016/1041-6080(92)90005-Y

Steffe, L. P. (1996). Radical constructivism a way of knowing and learning [Review of the same title, by Ernst von Glasersfeld]. *Zentralblatt Für Didaktik Der Mathematik [International Reviews on Mathematical Education]*, *96*(6), 202–204.

Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. *Handbook of Research Design in Mathematics and Science Education*, 267–306.

Thompson, P. W. (1993). Quantitative reasoning, complexity, and additive structures. *Educational Studies in Mathematics*, *25*(3), 165–208.

Thompson, P. W. (1994). Images of rate and operational understanding of the fundamental theorem of calculus. *Educational Studies in Mathematics*, *26*(2), 229–274.

Thompson, P. W. (1996). Imagery and the development of mathematical reasoning. In L. P. Steffe, P. Nesher, P. Cobb, G. A. Goldin, & B. Greer (Eds.), *Theories of mathematical learning* (pp. 267–283). Hillsdale, NJ: Erlbaum.

Thompson, P. W. (2000). Radical constructivism: Reflections and directions. In L. P. Steffe & P. W. Thompson (Eds.), *Radical constructivism in action: Building on the pioneering work of Ernst von Glasersfeld* (pp. 412–448). London: Falmer Press.

Thompson, P. W. (2008). Conceptual analysis of mathematical ideas: Some spadework at the foundation of mathematics education. In *Proceedings of the Annual Meeting of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 45–64). Retrieved from http://bit.ly/10YE9al

Thompson, P. W. (2011). Quantitative reasoning and mathematical modeling. In L. L. Hatfield, S. Chamberlain, & S. Belbase (Eds.), *New perspectives and directions for collaborative research in mathematics education* (Vol. 1, pp. 33–57). Laramie, WY: University of Wyoming. Retrieved from http://bit.ly/15II27f

Thompson, P. W. (2013). In the absence of meaning... In K. Leatham, *Vital Directions for Research in Mathematics Education* (pp. 57–93). New York, NY: Springer. Retrieved from http://bit.ly/Ztg3Hm

Thompson, P. W. (2016). Researching mathematical meanings for teaching. In L. D. English & D. Kirshner (Eds.), *Third Handbook of International Research in Mathematics Education* (pp. 435–461). New York: Taylor and Francis.

Thompson, P. W., Carlson, M. P., Byerley, C., & Hatfield, N. J. (2014). Schemes for thinking with magnitudes: A hypothesis about foundational reasoning abilities in algebra. In L. P. Steffe, L. L. Hatfield, & K. C. Moore (Eds.), *Epistemic algebra* *students: Emerging models of students' algebraic knowing* (pp. 1–24). Laramie, WY: University of Wyoming. Retrieved from http://bit.ly/1aNquwz

Thompson, P. W., & Silverman, J. (2008). The concept of accumulation in calculus. In M. P. Carlson & C. Rasmussen (Eds.), *Making the connection: Research and teaching in undergraduate mathematics* (Vol. 73, pp. 43–52). Washington, DC: Mathematical Association of America. Retrieved from http://bit.ly/Zth3LD

Tintle, N., Chance, B. L., Cobb, G. W., Rossman, A. J., Roy, S., Swanson, T., & VanderStoep, J. (2016). *Introduction to statistical investigations*. John Wiley & Sons, Inc.

Tufte, E. R. (2006). *Beautiful evidence*. Cheshire, Conn: Graphics Press.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the international congress of mathematicians* (Vol. 2, pp. 523–531). Retrieved from http://www.mathunion.org/ICM/ICM1974.2/Main/icm1974.2.0523.0532.ocr.pdf

von Glasersfeld, E. (1987). Preliminaries to any theory of representation. In C. Janvier (Ed.), *Problems of Representations in the Teaching and Learning of Mathematics* (pp. 215–225). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

von Glasersfeld, E. (1995). Radical constructivism: A way of knowing and learning. Routledge.

von Glasersfeld, E. (2001). Scheme theory as a key to the learning paradox. In A. Tryphon & Vonèche, J. Jacques (Eds.), *Working with Piaget: Essays in Honour of Bärbel Inhelder* (pp. 141–148). Psychology Press.

von Mises, R. (1981). *Probability, statistics, and truth*. New York: Dover Publications.

Wagenaar, W. (1991). Randomness and randomizers: Maybe the problem is not so big. *Journal of Behavioral Decision Making*, *4*(3), 220–222.

Watson, J. M. (2009). The influence of variation and expectation on the developing awareness of distribution. *Statistics Education Research Journal*, *8*(1), 32–61.

Weisberg, H. I. (2014). *Willful ignorance: the mismeasure of uncertainty*. Hoboken, New Jersey: Wiley.

Wild, C. J. (2006). The concept of distribution. *Statistics Education Research Journal*, *5*(2), 10–26.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review / Revue Internationale de Statistique*, *67*(3), 223. https://doi.org/10.2307/1403699

Wilson, M. (2005). *Constructing measures : an item response modeling approach*. Mahwah N.J.: Lawrence Erlbaum Associates.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching*, *46*(6), 716–730. https://doi.org/10.1002/tea.20318

Wilson, M. (2012). Responding to a challenge that learning progressions pose to measurement practice: Hypothesized links between dimensions of the outcome progression. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning progressions in science: Current challenges and future directions* (pp. 317–344). Boston: Sense Publishers.

APPENDIX A

CLINICIAL INTERVIEW QUESTIONS

The following are example questions to be used during the initial clinical interview.

Questions and tasks from and/or based on those in the literature have references listed.

How would you explain "probability" to another person? (from N. J. Hatfield, 2016a, 2016b)

What comes to your mind when you hear the term "variable"?

What comes to your mind when you hear the term "distribution"?

a) Select all of the following situations that you believe match your meaning for "random".
b) Tom and Harry are in the break room discussing what they thought about Star Wars: The Force Awakens. While describing what he liked about the movie, Tom said "Oh, did you know that Linda (a co-worker) is Lutheran?" Harry replied, "That's random."
c) You're at home, someone knocks on your door and you don't know who it is.
d) You and your two closest friends are trying to resolve who gets to choose what movie to see. One friend doesn't care but the other one and you both want to go see different movies. The neutral friend picks a number at random and the closest person wins.
e) Nothing is ever random; there is always a reason that things occur.
f) Everything is random.
g) A sequence is random when you can't find a pattern to it; like the number pi.
h) A sequence is random when you can't find a pattern, but you can use it predict something in the long-run.
i) None of these match my meaning for "random".

Which one of the following most closely matches your meaning for "probability"?
a) Probability is the ratio (fraction) of how many ways you can get a particular outcome compared to the total number of outcomes.
b) Probability is the likelihood that you have for observing some event.
c) Probability is the long-run relative frequency of observing some event.
d) Probability is the chance that you have for observing some event.
e) Probability is how often you see some event occur in a set of observations.

Suppose you toss a coin 20 times and get 19 heads and one tail. If you toss the coin one more time, what is the probability that coin will land heads?

What is the probability that ASU will win its next football game against UA?

Explain as though to someone who does not know what "probability" means what the number "3/36" represents in the following statement: The probability of getting a 4 when you roll two dice and multiply their face values is 3/36. (from N. J. Hatfield, 2016a, 2016b)

Four machines designed to mimic a fair coin produced the given sequences. Select the machines that you believe model a fair coin correctly.
Machine 1: H T H T H T H T H T H T H T H T H T H T
Machine 2: H H H H H H H H H H T T T T T T T T T T
Machine 3: H H T H H T H H T H H T H H T H H T H H
Machine 4: H T T H H T T T T H T T H T T T T H T T (from Green, 1987 as listed in Falk and Konold, 1994)

How would you explain "chance" to another person?

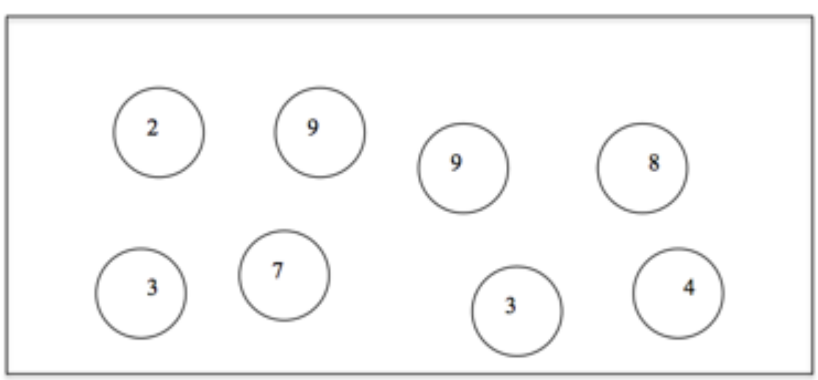
How would you explain the idea of "randomness"?

Explain as though to someone who does not know what "probability" means what the number "1/12" represents in the following statement: The probability of getting a 4 when you roll two dice and multiply their face values is 1/12.

What does it mean when a weather forecaster says that tomorrow there is a 70% chance of rain? What does the number, in this case 70%, tell you? How do they arrive at a specific number? (from Konold, 1989)

Suppose the forecaster said that there was a 70% chance of rain tomorrow and, in fact, it didn't rain. What would you conclude about the statement that there was a 70% chance of rain? (from Konold, 1989)

What do you think a random variable is?



Using the Lotto Box answer the following question. Suppose you choose a ball from the box at random. What is the probability that the ball will have a number that is 5 or larger?

Which one of the following most closely matches your meaning for "chance"?

a) Chance is the ratio (fraction) of how many ways you can get a particular outcome compared to the total number of outcomes.
b) Chance is the long-run relative frequency of observing some event.
c) Chance is how often you see some event occur in a set of observations.
d) Chance is the likelihood that you have for observing some event.
e) Chance is the probability that you have for observing some event.

Suppose that a baby was just born in your local hospital. What is the probability that the baby is a girl? Please explain your reasoning.

What do you think is meant by the phrase "distribution of a random variable"?

Explain as though to someone who does not know what "probability" means what the number "3/36" represents in the following statement: The probability of randomly selecting a US man, age 20-45, whose height is greater than 6ft is 3/36.

How would you explain "likelihood" to another person?

Which one of the following most closely matches your meaning for "likelihood"?

a) Likelihood is the long-run relative frequency of observing some event.
b) Likelihood is the probability that you have for observing some event.
c) Likelihood is the chance that you have for observing some event.
d) Likelihood is how often you see some event occur in a set of observations.
e) Likelihood is the ratio (fraction) of how many ways you can get a particular outcome compared to the total number of outcomes.

How would you explain "odds" to another person?

Consider the words "probability", "chance", and "likelihood". Which of these words do you believe refer to the same idea?

Suppose you hear someone state that among young adults (25 to 34 years old), the probability of a person having only a high school diploma is 0.31.

A) Rank each of the following statements (1-lowest, 5-highest) based upon how useful the statement is for explaining the idea of probability in this context.
    a. A probability of 0.31 means that 31/100ths of the sampled young adults will only have a high school diploma.
    b. A probability of 0.31 means that we will have exactly 31 young adults who only have a high school diploma for every 100 young adults sampled.
    c. A probability of 0.31 means that 31% of the time we repeat the sampling process to select a young adult, we'll observe a young adult who only has a high school diploma.
    d. A probability of 0.31 means that we have a 31% chance of selecting a

young adult who only has a high school diploma on the next trial of the sampling process.

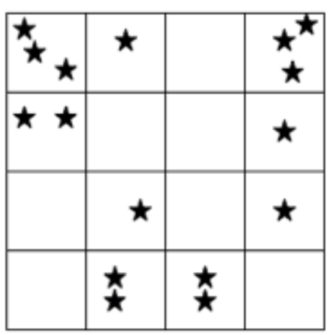
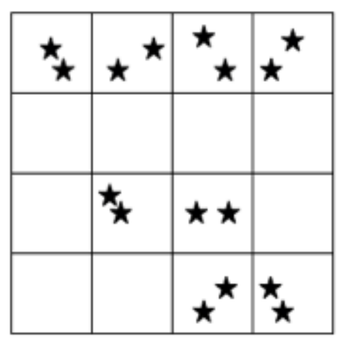
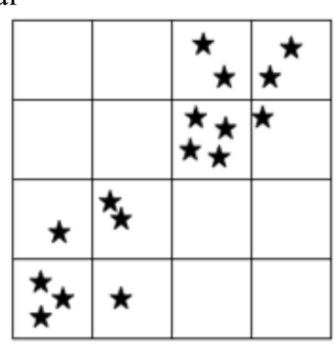
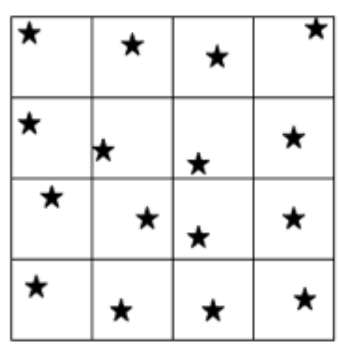
e. A probability of 0.31 means that out of 100 different categorizations of education levels, there are 31 that involve only a high school diploma.

B) Explain your reasoning for your rankings.

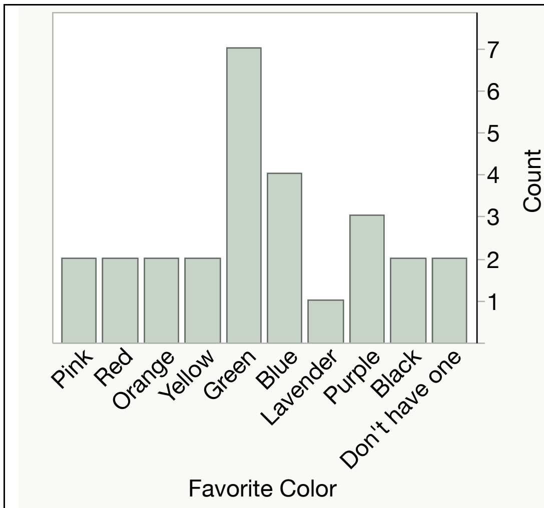
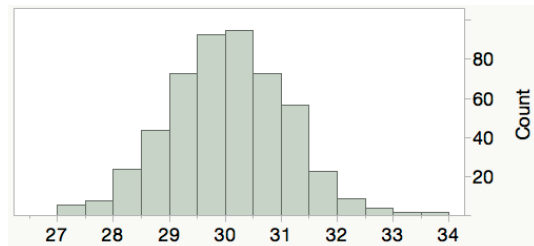
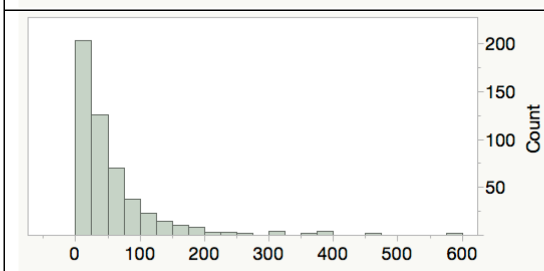
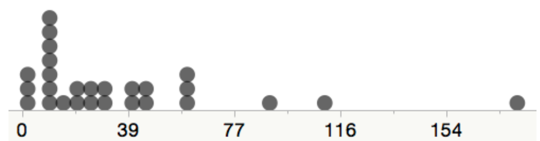
Imagine a game where you place the numbers 1 through 16 on slips of paper into a bag. You shake the bag and draw out one slip of paper. You then put a marker in the appropriate box that matches the number you drew out of the hat. You return the slip of paper to the bag and shake. You do this several times.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Four students (Alex, Beryl, Cristobal, and Dora) played this game. Examine their game boards in turn and determine whether or not the student cheated (i.e., didn't follow the game rules). For each student state whether or not the student cheated and explain how you decided. (adapted from Batanero & Serrano, 1999)

Alex 	Beryl 
Cristobal 	Dora 

Select any of the following options that you believe exemplifies the concept of distribution in Statistics. You may select as many as you feel is appropriate.

	
	
A distribution is a visual arrangement of data.	The distribution of a variable tells us what values the variable takes and how often the variable takes these values.
The distribution of a data set is a table, graph, or formula that provides the values of the observations and how often they occur.	The entire collection.
Any complete description of the behavior of a variable.	$C_D(x) = P\left[X \le x \mid X \sim D(\theta)\right]$ for all possible x of X.

If none of these options exemplify the concept of distribution, please provide how you would explain this concept.

Of the options you selected, which one do you feel is the closest/best match to how you think about distribution? Why?