

CONCEPTUAL ISSUES IN QUANTIFYING UNUSUALNESS AND CONCEIVING STOCHASTIC EXPERIMENTS: INSIGHTS FROM STUDENTS' EXPERIENCES IN DESIGNING SAMPLING SIMULATIONS

LUIS SALDANHA
Université du Québec à Montréal
saldanha.luis@uqam.ca

ABSTRACT

This article reports on a classroom teaching experiment that engaged a group of high school students in designing sampling simulations within a computer microworld. The simulation-design activities aimed to foster students' abilities to conceive of contextual situations as stochastic experiments, and to engage them with the logic of hypothesis testing. This scheme of ideas involves imagining a population and a sample drawn from it, and an image of repeated sampling as a basis for quantifying a sampling outcome's unusualness in terms of long-run relative frequency under an assumption about the population's composition. The study highlights challenges that students experienced, and sheds light on aspects of conceiving stochastic experiments and conceiving a sampling outcome's unusualness as a probabilistic quantity.

Keywords: *Statistics education research; Statistical quantity; Relative frequency*

1. INTRODUCTION AND BACKGROUND

In educational settings the use of simulation has been advocated as a potentially useful pedagogical tool to help students develop meaning for the concepts of probability, inference, hypothesis testing, and related statistical ideas (Jones, Langrall, & Mooney, 2007; Konold, 2002; Mills, 2002; NCTM, 2000; Shaughnessy, 1992, 2007; Stohl & Tarr, 2002). Indeed, recent developments in introductory statistics curricula centralize the use of simulation software as a key innovation (Garfield, delMas, & Zieffler, 2012; Rossman & Chance, 2012). This is also a very active area in statistics education research, as indicated by the recent proliferation of articles presented at research sessions of the ICOTS9 conference of 2014 (e.g., Chance & McGaughey, 2014; Tintle et al., 2014; Zieffler, delMas, Garfield, & Brown, 2014). But what is entailed conceptually in understanding and using simulations? Further, what challenges might students experience in their efforts to design and use simulations to make informal inferences and draw probabilistic conclusions about situations that involve conceiving a stochastic experiment? This article addresses these questions by reporting on part of a classroom teaching experiment that engaged a group of high school students in designing simulations within the *Prob Sim* (Konold & Miller, 1996) sampling and probability microworld—a precursor of the sampler tool now integrated into the *TinkerPlots*TM software (Konold & Miller, 2011).

1.1. PRIOR RELEVANT RESEARCH

A number of notable studies have explored students' statistical reasoning or achievement in relation to their participation in instruction involving the use of computer simulations (Biehler & Prömmel, 2010; Chance & McGaughey, 2014; delMas, Garfield, & Chance, 1999; Drier, 2000; Konold, Harradine, & Kazak, 2007; Kuhn, Hoppe, Lingnau, & Wichmann, 2006; Maxara & Biehler, 2006, 2007; Noll, Gebresenbet, & Glover, 2016; Pratt, 2000; Saldanha & Thompson, 2014, 2007; Sedlmeier, 1999; Stohl & Tarr, 2002; Tintle et al., 2014; Well, Pollatsek, & Boyce, 1990; Zieffler et al., 2014). Some of this research focused on students' use of specific software features when running simulations within a computer microworld, pointing to aspects of such use that shaped their conceptual development. A subset of those studies focused on issues pertaining to students' efforts to *design* simulations per se.

Aspects of using simulations Pratt (2000) explored the development of a pair of students' thinking as they engaged with tasks involving the use of the *Chance Maker* (CM) microworld. CM provided students with virtual spinners and dice whose internal workings they could manipulate to specify the number of trials of an experiment, and to produce desired distributions of outcomes displayed as pie graphs and dot plots. Tasks involved "mending" the devices by manipulating their internal workings in an effort to ensure that multiple trials of spinning two spinners or tossing two dice produced a distribution of outcomes of their sum that reflected "fair" spinners or dice. Pratt reported that such mending activity afforded opportunities for the students to notice that the resulting CM graphs produced uniform distributions whenever each possible outcome was listed once in the devices' workings boxes. The students consequently developed a conception of fairness that was connected to the equal frequency of different combinations of outcomes specified in the devices' workings boxes.

Stohl and Tarr (2002) studied the meaning-making processes of a pair of sixth grade students as they engaged with tasks involving use of the *Probability Explorer* (PE; Stohl, 2002) microworld to simulate experiments and analyze simulated data. PE provided students with virtual spinners and dice, enabling them to assign weights to spinners' colors, to specify the number of trials of a sampling experiment, and to choose graphical representations to display the results of simulated experiments. Students used these features to model pre-given stochastic situations (e.g., spinning a disk comprised of 3 colored sectors in a given proportion), and to draw connections between empirical and theoretical probabilities by comparing the latter with the distribution of outcomes produced by running their model in PE as a simulation. Students reportedly learned that inferences made on the basis of small samples often led to erroneous conclusions, whereas those based on sufficiently large samples tended to be reliable. Drier (2000) reported that fourth grade students used PE's graphical representations of data as objects of analysis in the context of exploring an "evening-out" phenomenon among sampling outcomes. Through experimentation with the software, students reportedly came to recognize that conducting more trials of a sampling experiment tended to produce distributions of sampling outcomes that closely resembled what they expected according to theoretical probabilities.

Kuhn et al. (2006) studied high school students' collaborative use of a networked "stochastic experiment" simulation environment that involved modeling and exploring the likelihood of certain outcomes in real world contexts (e.g., outcomes of a lottery). The simulation environment provided students with virtual tools such as dice and urns that enabled them to control parameters such as number of tosses or draws. Students'

experimentation with the software, and their comparisons of the relative frequency of outcomes of simulated dice tosses over sufficiently large numbers of trials, provoked some to formulate a naïve version of the law of large numbers.

A salient commonality among these studies was the use of simulation tools that were virtual versions of canonical random devices such as spinners and dice. These devices provided a high level of transparency to the modeled situations, essentially providing students with pre-determined models to use in the modeling process. Additionally, although each of these studies engaged students within a microworld that provided opportunities for them to easily experiment with, and control, various parameters such as the number of trials of a simulated experiment, the focus of the research was not on students' conceptualizations of an experiment. Konold et al. (2007) studied middle school students' use of a different simulation tool—a sampler within the *TinkerPlots 2.0* software (Konold & Miller, 2011)—to generate realistic data that met specified criteria. The authors reported that this created opportunities for students to problematize the concepts of a case, its attributes, and values thereof. Additionally, the sampler enabled students to efficiently run many trials of a simulated experiment, thereby reducing the length of the model creation and assessment cycle. Konold et al.'s (2007) report, however, did not focus on students' conceptualizations of the experiments they modeled, or what they understood they were modeling with the sampler. More recently Zieffler et al. (2014) reported on an interview study of first year university students' reasoning when modeling probabilistic experiments, after they had participated in an instructional unit focused on modeling and simulation also involving the use of *TinkerPlots'* sampler tool (Garfield et al., 2012). Their report highlighted that despite observing a diversity of levels of statistical reasoning among the students, the use of *TinkerPlots* in the context of interview activities seemed to support the emergence of a deeper reasoning about uncertainty among some students. Moreover, many students seemed to have internalised the *TinkerPlots* models as evidenced by their mimicking them in their explanations and vocabulary prior to being prompted to use the software in interview tasks. Details regarding students' actual construction of models and design of simulations were not, however, the focus of the study.

Aspects of designing simulations An explicit focus on aspects of designing simulations per se is seen in the work of Noll et al. (2016), and Biehler and colleagues (Biehler & Prömmel, 2010; Maxara & Biehler, 2006, 2007). Noll et al. documented the thinking of college students enrolled in a simulation-based introductory statistics course that engaged them in using *TinkerPlots'* sampler tool to design and run simulations that modeled situations involving the testing of statistical hypotheses. Despite observing substantial gains in the number of students who developed appropriate reasoning about statistical modeling and simulation, from pre- to post-instruction, Noll et al. reported that half of their students experienced persistent difficulties in building correct simulation models and running a single trial of the simulation. The authors identified the design of a model with the sampler tool as one of the students' biggest challenges. In particular, students had difficulty identifying the population to be sampled and its composition, and envisioning what constituted an appropriate sample and its statistic. Such problems suggest difficulties in conceiving of the experiment being modeled and of what constitutes a trial of an experiment.

Biehler and colleagues (Biehler & Prömmel, 2010; Maxara & Biehler, 2006, 2007) researched students' and teachers' uses of *Fathom* software (Finzer, 2012) in designing stochastic simulations. In one study (Maxara & Biehler, 2006) they reported that students tended to prematurely use the software before thinking through the stochastic aspects of

their simulation. Similarly, the pre-service teacher participants in a subsequent study (Maxara & Biehler, 2007) reportedly experienced difficulties in using Fathom successfully because they did not clearly distinguish random variables and events in the modeling and planning phase of their simulations, prior to using the software. The authors concluded that the planning phase of simulation-design activities should place greater emphasis on supporting teacher reflection about the relevant probabilistic ideas, and in particular should involve explicitly connecting probabilistic components of their stochastic models with the features and constraints of the Fathom environment. Biehler and Prömmel (2010) followed up on these implications in a subsequent study involving the comparison of two approaches to the planning of stochastic simulations on high school students' learning and simulation-design competencies. One group of students completed the plan of their simulation before working on the computer, whereas the other group integrated the planning and computer implementation activities. Although both approaches were reportedly equally successful in supporting students' simulation-design competencies, the group that separated planning and implementation exhibited significantly more uniform solution processes and required significantly less assistance in the computer implementation phase of the work.

Noll et al.'s (2016), Biehler and Prömmel's (2010), and Maxara and Biehler's (2006, 2007) research indicates that creating stochastic simulations can be a very challenging endeavor for learners, and suggests that it may be advantageous to have learners separate the planning of stochastic simulations from their implementation within the software being used. The research reported in the present article is in line with this suggestion; it attempts to gain insight into conceptual issues pertaining to the *design* of simulations within a microworld, before their implementation, and it takes a detailed view of students' experiences as they engaged in such design activity. In addition to being an area of statistics education about which relatively little is known, its importance is further underscored by the recent development, and increasing prevalence, of new statistics curricula that centralize the use of computer simulations in the teaching of inference (Garfield et al., 2012; Rossman & Chance, 2012).

2. PURPOSE

This article aims to shed light on conceptual operations and challenges entailed in construing contextual situations as stochastic experiments, and in conceiving of a random sampling outcome's anticipated unusualness as a statistical quantity. The article is based on a classroom teaching experiment which explored the thinking of a group of high school students that emerged as they engaged with tasks involving the design and use of computer simulations for making informal statistical inferences and testing hypotheses in situations involving the construal of a stochastic experiment. I write this article from the perspective of a member of the research team that conducted the teaching experiment, and as a witness of the instructional interactions that unfolded therein.

3. THEORETICAL FRAMING

Two theoretical perspectives shaped and framed both the design of instruction in the experiment, and my analysis of the data generated in it. I refer to the first perspective as *stochastic conception of sampling*. Drawing on von Mises' (1957) strong frequentist perspective of probability, and on Kahneman and Tversky's (1982) *distributional* perspective of sampling, this is a conception that entails a fundamental underlying image of sampling as a random selection process that can be repeated under essentially identical

conditions, and of each resulting individual sampling outcome (i.e., the value of a sample statistic) as one amongst a class of such outcomes generated by repeating the process. Liu and Thompson (2007) sharpened these important ideas by foregrounding the repeatability of the underlying generation process, in what they refer to as a stochastic conception of an event: “A person having a stochastic conception of an event conceives of an observed outcome as but one expression of an underlying repeatable process (what Horvath and Lehrer (1998) called a “trial”), which over the long run will produce a stable distribution of outcomes.” (Liu & Thompson, 2007, p. 122). Additionally, Liu and Thompson dissected a stochastic conception of an event into the following constitutive network of conceptual operations: conceiving of a random sampling process—that is, randomly selecting a sample of a given size from a population and determining the value of a statistic of interest for that sample; imagining repeating the random sampling process under similar conditions; anticipating that repeating the process would produce a collection of outcomes (i.e., values of the sample statistic); and understanding that although the random selection process produces variability in the outcomes, the distribution of outcomes will become stable over the long run (Liu & Thompson, 2007).

A stochastic conception of sampling was at the core of the instructional agenda of the teaching experiment discussed here; it was an explicit goal of instruction that students develop such a conception, and instructional tasks and classroom interactions around them aimed to foster its development. Accordingly, my analyses of the data generated in the experiment drew on this perspective as an interpretive lens and explanatory construct.

The second theoretical perspective that shaped this study is Thompson’s (1994) theory of quantitative reasoning, pertaining to the psychology of conceiving situations in terms of quantities and relationships among them. In Thompson’s theory a quantity is viewed as a conceptual entity that a person constructs when conceptualizing situations as having measurable attributes. This is described schematically: “[Quantity] involves an object-image, a conceptualized attribute of the object, a tacit understanding of appropriate units of measure, and a quantification process by which one directly or indirectly assigns numerical values to the attribute” (Cortina, Saldanha, & Thompson, 1999, p. 467). Also see Thompson (1994).

Thompson’s (1994) theory of quantitative reasoning shaped this study in two important and interrelated ways. First, the theory frames the kind of reasoning targeted with regard to the statistical ideas addressed in instruction, thus orienting the design of particular instructional activities aimed at fostering the development of such reasoning. Specifically, the ideas of sample statistic and population parameter were framed in terms of measurement and quantity: Numerical data were presented and discussed in instruction as measures of sampling outcomes (e.g., measures of an attribute of a collection of items selected from a population). Similarly, a population parameter was cast as a measure of an attribute of the population of sampled items. Instruction engaged students in working with collections of such data values in ways intended to foster their ability to quantify attributes of them; in particular, to quantify a sampling outcome’s likelihood or unusualness in terms of the outcome’s expected long-run relative frequency. In the remainder of the paper I will often refer to this latter idea as *quantifying unusualness* or *unusualness as a statistical quantity*.

A second way in which Thompson’s (1994) theory is germane to this study is that it is a cognitive-based theory directed toward explicating conceptual operations (von Glasersfeld, 1995)—mental images, ways of thinking, and coordinations thereof—by which people come to conceive situations quantitatively. The frame of conceptual operations oriented my analysis of students’ reasoning about sampling as a stochastic event and quantifying their expectation of sampling outcomes. Specifically, this frame

provided a useful reference point for describing and explaining students' reasoning about those ideas, and for my choice of appropriate grain size of such descriptions and explanations.

To conclude I emphasize the key idea in Thompson's (1994) theory that conceiving of quantity entails *conceptualizing* given situations in ways that support thinking of attributes embedded within them as measurable. This deeply constructivist idea is expressed in this particular study by my stance that a sampling event is not *in and of itself stochastic*, in an ontological sense. Rather, an event must be *construed by a person* in such a way as to support conceiving of the likelihood of its occurrence and its anticipated unusualness as statistical quantities. From this perspective, the question of what conceptual operations and challenges might be entailed in developing such a construal becomes an important focus of inquiry.

4. METHOD

4.1. PARTICIPANTS

A sample of convenience consisting of eight academically diverse students participated in a multi-phase classroom teaching experiment conducted within an intact introductory statistics course at a suburban high school in the Southeastern United States. Prior to their participation in the experiment, all students had completed a standard Algebra II course that included a short unit on statistics and probability. Examination of the relevant textbook unit suggested that students had not received instruction in the statistical ways of thinking targeted in the teaching experiment (elaborated in the next section). A pre-assessment querying students' intuitions and understandings of sampling and inference suggested that they had largely informal and non-stochastic understandings of these ideas prior to participating in the teaching experiment. Most understood a sample to be akin to a little part of something (Watson & Moritz, 2000), and their ideas of variability were focused on differences in outcomes without entailing ideas of repeated sampling or distribution. In the next section I summarize the repeated sampling experiences students had before the phase of the experiment reported in this article.

4.2. INSTRUCTIONAL SETTING AND ACTIVITIES

Instructional setting and style Instruction, which was conducted by the lead member of the research team, employed activities designed and enacted as discussion-based and inquiry-oriented investigations. Instruction emphasized describing and explaining statistical ideas and connections among them over the use of formal symbols and calculations. The research team negotiated a culture of sense-making in the classroom by placing a high premium on, and promoting, active participation such as listening, reflecting, questioning, conjecturing, and explaining and describing one's own and others' thinking about the statistical ideas under discussion.

Given the team's dual agenda of conducting research within an instructional setting, the instructor employed an atypical teaching style consistent with Steffe and Thompson's (2000) idea of *teaching as a method of scientific investigation*. In this style, teaching is employed as a tool of inquiry into students' reasoning and conceptions about particular mathematical ideas; it is used to influence students' thinking as much as to generate data that serve as a basis for modeling their thinking, both extemporaneously and retroactively. The style is consequently marked by particular teacher actions that could be misinterpreted as simply "telling," whereas the intent is to assess how particular

statements and ideas might be assimilated by students and what accommodations they are able to make in trying to understand them. Such actions include regularly probing students' thinking, inciting them to engage in repeated reasoning (Harel & Koichu, 2010) and scaffolding their efforts to do so, and pressing them to articulate such reasoning.

Activities and microworld In the phase of the teaching experiment immediately preceding the one reported here, students had participated in a sequence of lessons that engaged them with both tactile and simulated sampling activities. Those activities involved repeating the experiment of selecting a small random sample of objects from a dichotomous population, generating values of a sample statistic (e.g., the number of red candies in a sample), and creating frequency distributions of those values. Students used such distributions as a basis for inferring the underlying population's composition, and for quantifying how unusual a particular sampling outcome was in terms of its observed relative frequency. In that context students developed the idea that a particular value of the sample statistic would be judged as “unusual” if it occurred infrequently in the distribution of the statistic (Saldanha & Thompson, 2007).

In the subsequent phase of the experiment—the basis for the present article—students engaged in activities involving the design of sampling simulations within the Prob Sim microworld (Konold & Miller, 1996). Figure 1 displays the Prob Sim user interface.

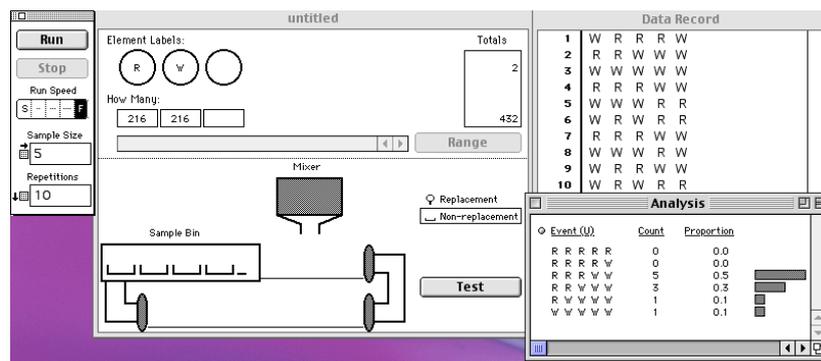


Figure 1. The Prob Sim user interface, data display, and analysis windows

Prob Sim employed the metaphor of a “mixer” for a population, and a “bin” for a sample. It enabled the user to easily specify a population's composition and size, the size and selection method of a sample (with or without replacement), and the number of trials of the simulated sampling experiment to be conducted (Figure 1, left side windows). Moreover, the software displayed the outcome of each simulated sampling experiment as a raw data list. Finally, Prob Sim provided summary analyses of the aggregated outcomes, and displayed their distribution as a relative frequency histogram (Figure 1, right side windows).

The simulation design activities asked students to explore questions of the type “is event x unusual?” Students were presented with scenarios in textual form that described a given situation in non-statistical terms, and that framed such questions for them to explore. Their task was to design a simulation, guided by the constraints of the Prob Sim interface, to run the simulation, and then to draw a conclusion on the basis of its results. Figure 2 displays an example of a simulation design activity discussed later in this report.

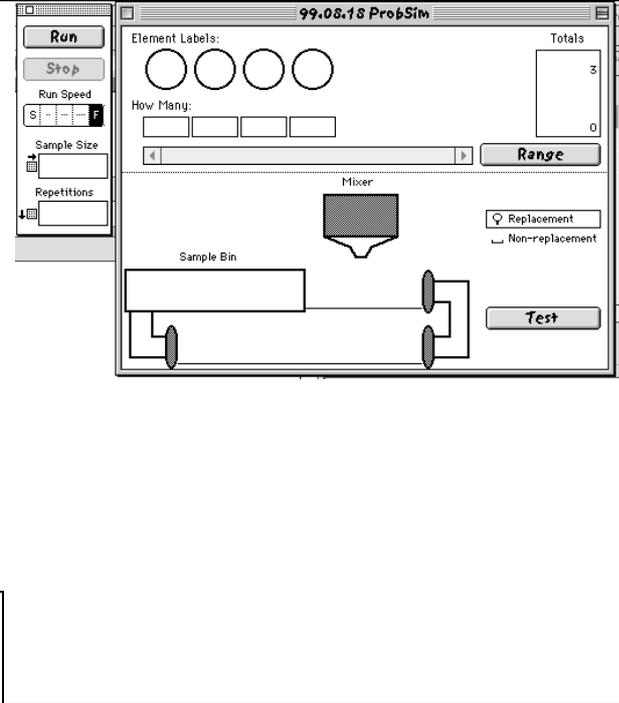
<p>Investigating “Unusualness”</p> <p>Ephram works at a theatre, taking tickets for one movie per night at a theatre that holds 250 people. The town has 30,000 people. He estimates that he knows 300 of them by name.</p> <p>Ephram noticed that he often saw at least two people he knew. Is it in fact unusual that at least two people Ephram knows attend the movie he shows, or could people be coming because he is there? (The theatre holds 250 people)</p> <ol style="list-style-type: none"> 1. Assumptions for your investigation: 2. Method of investigation: 3. Result: <table border="1" data-bbox="443 853 708 920"> <tr> <td>“Gut level” answer:</td> </tr> </table> 4. Conclusion: 	“Gut level” answer:	
“Gut level” answer:		

Figure 2. A task involving the design of a sampling simulation (left panel) and the use of Prob Sim (right panel) to investigate the issue of whether a particular outcome is unusual (adapted from Konold, 2002)

The activities shared a common feature and structure: Students were presented with a contextual situation that involved investigating whether there was reason to believe that a specified event of interest was unusual, and they were provoked to structure their investigation according to the guidelines suggested by the four prompts shown in Figure 2. The simulation design activities unfolded over four 45-55 minute class periods held on consecutive school days. These sessions involved whole-class discussions directed at having students think and describe how they would use Prob Sim to investigate and resolve the question raised in a given situation.

Intent of the activities An important clarification is warranted here regarding the intent of these activities: The overarching instructional goal was *not* that students would become skilled at designing such simulations per se. Rather, the primary aim was to provide students with occasion to construe contextual situations as idealized probabilistic experiments, and to thus provoke them to grapple with whatever complexities might be entailed in doing so. The research team reasoned that such a construal involves conceptualizing a scheme composed of the following key ideas: an imagined population and sample to be drawn from it, a method of selection, and repeated sampling and its aggregation into a distribution of outcomes as a method for exploring the relative unusualness of a particular sampling outcome under an assumption about the population’s composition. As such, the activities were designed to promote students’ conceiving of probabilistic situations (Konold, 1989), sampling as a stochastic event, and to engage them with the logic of statistical hypothesis testing. It is important that the reader keep in mind the intended goal of these activities when considering the analyses and results of the

study, as those focus only on issues that emerged in students' efforts at design, and *not* on issues pertaining to the actual implementation or running of the simulations.

I must also emphasize that the research team had already devoted considerable effort to developing the above-mentioned key ideas with students in the preceding phase of the experiment (Saldanha & Thompson, 2007). However, in contrast to the sampling activities in that phase, the simulation design activities in the current phase aimed to extend students' abilities to conceptualize probabilistic experiments in situations that were not explicitly cast in statistical terms or that mentioned canonical objects (e.g., cards or coins) which could serve as surface cues for how to proceed. To help students transition to the simulation design activities using Prob Sim, they were preceded by discussions around demonstrations of the microworld that established the concepts of random sample and population, making an inference from the former to the latter, and sampling with and without replacement. Students subsequently spent some time experimenting with the microworld in pairs.

4.3. DATA CORPUS AND ANALYSIS

This report draws principally on the classroom discussions that unfolded around the simulation design activities, to illustrate issues that students grappled with during their efforts to design the simulations. These discussions were captured on audio-video recordings of the classroom sessions and then transcribed for detailed analysis. Additional evidence is drawn from students' written responses to a post-experiment test question that queried their thinking about a key idea addressed in the simulation design activities.

Analysis of the classroom discussions began with a chronological viewing of the lesson videos, resulting in the creation of an outline of each lesson and the activity sequence across them, and in the identification of episodes that evidenced apparent challenges and difficulties experienced by students (Powell, Francisco, & Maher, 2003). These flagged episodes were examined in detail in a second phase of analysis, wherein I generated initial interpretations of them. This entailed describing the activity in which each episode arose, characterizing the content of the discussions that emerged therein, with a focus on the mathematical thinking that students expressed in those discussions, including challenges and difficulties that they seemed to experience. In this second phase, I employed quantitative reasoning theory (Thompson, 1994) as an interpretive lens; discussions were analyzed for whether, and to what extent, students were thinking about the activity situations in ways consistent with a stochastic conception of an event and therefore amenable to conceiving of a sampling outcome's anticipated unusualness in terms of long-run relative frequency. In particular, I characterized the focus of students' images, mental actions, and coordination thereof, as pointers to the ways in which they might or might not have been composing and connecting ideas. In a third phase of analysis I examined these characterizations in chronological order for prevalent or recurrent themes across them, with a particular interest in whether the challenges and difficulties identified across episodes were transient or persistent and arguably robust. This across-episode analysis was especially important given the nature of the classroom discussions; students' utterances and contributions within individual discussions were often tentative or halting, and heavily guided or eclipsed by the instructor's ideas and attempts to scaffold students' thinking. It was in the aggregate, by examining multiple discussions across the sequence of activities, that I was able to identify thematic issues and difficulties experienced by students.

The post-test question queried students' abilities to operationalize "unusual event" in quantitative terms. Students' responses were coded for whether they expressed key components of an understanding that are largely consistent or inconsistent with the understanding targeted in instruction. This coding was supplemented by more fine-grained analysis of the responses that characterized how their key features differed from the targeted conception and what those suggested about students' underlying images and coordination thereof.

5. RESULTS

The study's findings are framed in terms of two central themes that emerged from my analysis of students' engagement in the simulation design activities: 1) conceiving a sampling outcome's anticipated unusualness as a statistical quantity, and 2) construing a contextual situation as an idealized stochastic experiment. Although these themes point to interrelated conceptual issues, they are discussed separately for the sake of clarity. These themes are illustrated primarily with evidence drawn from the classroom discussions, and secondarily with the students' post-test responses.

5.1. THEME 1: QUANTIFYING UNUSUALNESS

The first situation with which students engaged was adapted from the so-called "birthday problem" (Figure 3) (Konold, 2002):

Jill was in a class of 25 people. The teacher asked each student for his or her birthday to enter in the class calendar. The class was surprised when two people had the same birthday! Is this, in fact, an unusual event?

Figure 3. The birthday problem

A discussion about this situation emerged out of the instructor's effort to solicit students' ideas about the meaning of *unusual event*; the synonyms "surprising" or "unexpected" were widely offered by students in response. The following discussion excerpt starts from the point immediately after students offered these interpretations; it indicates students' lingering thinking with regard to this idea.

Excerpt 1 (Lesson 5)

1. Intr: [...] So what does it mean for an event to be unusual?
2. Peter: You're not expecting it.
3. Kit: You don't expect it.
4. Instr: Yeah, it's unexpected, that in a large number of times that you do this (2-second pause) you, you, you expect to see it rarely. Ok? (2-second pause) All right, so in this particular case what does it mean that, to wonder if the event that's described is unusual?
5. Peter: You wouldn't expect 2 people in a cla—in a group of 25 people to have the same birthday.
6. Instr: Ok. Now, you're leaving out something. That's good as far as you went, but you're leaving something out [...] Are we talking, when we talk about an event being unusual, are we talking about just one occurrence? (4-second silence)
7. Instr: Are we just talking about that one class?
8. Luke: No. It's unusual—as many times as you did this test with the class, it would be unusual for 2 people to have the same birthday.

9. Instr: So you wouldn't expect it to happen very often.
10. Luke: Correct.
11. Instr: But the idea, the part that you, uhh both of you left out, uhh Peter and Kit, was if you looked at a whole bunch of classes, of size 25, you wouldn't find very many. You see? You left out the part of looking at a whole bunch of classes. All right? And (2-second pause), and that's, that's, that's a key idea. The idea that what we're talking about is doing something a large number of times, looking at a class of 25 students. Ok?

Students' descriptions generally suggest that their ideas about unusualness were not of a quantitative nature. Their images were focused on individual occurrences, and not embedded within an imagined sequence of trials of a repeatable sampling experiment that might support quantifying likelihood or unusualness in terms of long-run relative frequency. Indeed, ideas of (relative) frequency were still largely absent from students' thinking and discourse. It was the instructor who oriented the class to such quantification by explicitly pointing out what idea students were leaving out of their thinking, by couching unusualness in terms of infrequency and re-casting the situation in terms of a repetitive process (see line 11). Luke's response (line 8) is arguably exceptional, as it suggests that the idea of frequency was part of his conception of unusualness. However, his was still a very implicit image of frequency ("as many times as you did this test"). Luke's allusion to the idea of "testing" something appears to have been rooted in his image of repeating some process and determining whether the event of interest, "two people with the same birthday," was observed. However, Luke did not articulate what that process might be.

In addition to illustrating aspects of students' thinking at that juncture of the experiment, Excerpt 1 also illustrates the type of thinking targeted in instruction, as articulated by the instructor's utterances. The instructor's characterization in the above excerpt (line 11) aimed to build on students' experiences with repeated sampling gained in the preceding phase of the teaching experiment.

Excerpt 2 (following) is from the opening discussion of the subsequent day's lesson, in which the instructor aimed to recap the idea of framing unusualness in quantitative terms, as a long-run relative frequency. The excerpt illustrates both the robustness of students' non-quantitative conceptions of unusualness, and its resistance to an easy "repair" in the form of characterizations provided by the instructor in the previous day's discussion.

Excerpt 2 (Lesson 6)

1. Instr: Uhh, in last class we looked at how, how, ways in which we could investigate whether or not something was unusual. All right? Now, what did it m—what did we mean by saying that something was unusual?
2. Luke: It happens less than 50% of the time.
3. Instr: Or uhh, was that it? I mean, if it happened 49% of, like, is it unusual to get a tail if we toss a head?
4. Michelle: Unexpected
5. Luke: Yeah, it's unexpected.
6. Instr: All right, it's unexpected. And how, how would you quantify that?
7. Michelle: How would we what?
8. Instr: How would you quantify that, that it's unexpected?
9. Nicole: I mean—
10. David: It doesn't usually happen.
11. Instr: That's not quantifying it. That's putting it—what, what does that mean?
12. Luke: Rephrase it, David

Excerpt 2 illustrates students' continuing disinclination to think of unusualness in quantitative terms. Luke's response (line 2) seems exceptional, in that it apparently entailed the idea of relative frequency. However, I note an important subtle distinction: His focus was on a numerical criterion (50%) for deciding whether an event is unusual, rather than the *process* by which one might arrive at such a criterion. As such, from an expert observer's perspective, Luke responded as if the question had been "below what cut-off proportion do we call an event 'unusual'?" rather than "how can we think about an event so that we can determine how 'unusual' it might be?" This suggests that Luke did not yet possess the latter way of thinking, which would explain difficulties in conceptualizing the underlying situation in terms that support his conceiving unusualness quantitatively.

Responses like David's "it doesn't usually happen" (line 10) amount to a non-quantitative rephrasing of "it's unexpected," and were quite common in the earlier phases of the discussions. In a sense, the students and the instructor were speaking different languages; the students were living largely in the realm of feeling and intuition, whereas the instructor was attempting to nudge them into a quantitative mindset by provoking them to map those intuitions onto a relative frequency perspective. Excerpts 1 and 2 naturally raise the possibility that students' difficulties were due to a miscommunication; students may have simply been unaware of, or not understood, the instructor's exacting requirement for expressing unusualness in quantitative terms, and were therefore playing a different game. Although this possibility cannot be discounted, it is inconsistent with the fact that students had, in the preceding phase of the teaching experiment, already used the idea of repeated sampling as a basis for quantifying a sampling outcome's unusualness in terms of long-run relative frequency. A more plausible explanation is that students were not bringing their prior ways of thinking to the current situations, which were cast in non-statistical terms, and which lacked any explicit cues to do so.

Generating and composing requisite images The next two data excerpts are contiguous segments drawn from the classroom discussions around the "movie theatre scenario" (Konold, 2002) displayed in Figure 2. The excerpts illustrate both the instructional challenge of helping the students assume the quantitative perspective, and the tenuous and fragile nature of their efforts to negotiate a transition to it. More specifically, the excerpts highlight two particular aspects that were problematic for students in the quantification process: namely, generating and coordinating images necessary for quantifying unusualness as a relative frequency. Excerpt 3 highlights Kit's thinking.

Excerpt 3 (Lesson 6)

400. Instr: Let's all make sure that we know what's going on. What is, what is it that's at issue?
(7-second silence)
401. Instr: Kit?
402. Kit: Whether or not it's unusual for him to see at least 2 people that he knows.
403. Instr: And what does it mean, what does "unusual" mean?
404. Kit: Not expected.
405. Instr: Ok. Go on and quantify that.
(5-second silence)
406. Instr: It means if he were, if he were to do this many many times he would expect some small fraction of the time for this to happen, to see—see, keep, you gotta, I want you to keep putting this idea of repeating an event over and over and over again. (3-second pause). Ok, it's not a matter of feeling like "gosh, I don't

expect this to happen.” That’s not where likelihood is determined! Likelihood is determined in the actual repeating. Not your feeling about it, but rather the repetition of the event and the fraction of the time that something happens in those repetitions. Ok? So now, Kit, once more: what is, what does it mean to be unexpected?

407. Kit: Uhh, a small fraction of the time, when it’s done several times.
 408. Instr: Ok, and what is the “it” in this case?
 409. Kit: Uhh, the people—uhh seeing more than 2 people or 2 people one time per night, that he knows
 410. Instr: Ok, so over many, many nights, assuming he’s there just once a night, over many, many, many nights, we have a small fraction of those nights where he sees 2 or more people that he knows. Now, do you see how that quantifies, brings, brings ideas of quantity into it? It’s no longer just a feeling that he has about expectation. It’s rather, we’re talking about repeating something many, many times and just looking at the fraction of the time that something happens! (3-second pause) Now you’ll, it’ll—it’ll get so that this is second nature to you to start thinking this way. And you’ll also wonder how you could’ve thought any other way. But it takes practice, and I, that’s why I keep insisting that you bring this idea out in the open.

Kit evidently understood the issue to be whether the event of interest—seeing two or more acquaintances in the movie theatre—is unusual. However, her sense of unusualness seemed non-quantitative, as indicated by her utterance in line 404 and her lack of a response to the instructor’s prompt to quantify “unexpected” (line 405). The instructor then highlighted the difference between having a mere feeling that something is unexpected and thinking about expectation quantitatively, in terms of an imagined process that could be repeated many times and then considering the relative frequency with which the event of interest is observed (line 406). Kit’s subsequent attempt to describe “unexpected” in quantitative terms (lines 407 and 409) points to important aspects of her thinking. On the one hand, it indicates the prominence of the idea of “a small fraction of the time,” and that she might therefore have been mindful of an underlying repeatable process. On the other hand, Kit’s description is tenuous, and suggests an ambiguity in her thinking with regard to distinguishing the process being repeated and observing the event of interest; her response to the instructor’s call for clarification of the referent for “it” (lines 407-409) indicates that she imagined that observing two (or more) acquaintances in the movie theatre was itself the process being repeated. Thus, Kit seemed mindful of two images—repeating a process and observing the outcome of interest—but she was not apparently coordinating them to imagine this scenario: *Repeat the process of looking in the movie theatre once per night, on many nights, and keep track of the number of such repetitions together with the number of nights in which you recognize at least two people.*

The significance of Kit’s ambiguous referent for “it” in Excerpt 3 is that it highlights the important issue of the coordination of images necessary to construe the given scenario in a manner that could support quantifying the event’s likelihood or unusualness. As reflected by the instructor’s description (line 410), this construal can be seen to involve a two-image scheme, each entailing a sense of something accumulating: One image entails accumulating the number of repetitions of the observation process (i.e., looking in the theatre each night), whereas the other entails accumulating the number of times the event of interest “two or more acquaintances” is observed. Conceiving the event’s likelihood as a quantity entails distinguishing and coordinating the two images, keeping mental track of each accumulation, and comparing that of the second to the first in relative terms.

Difficulty in generating and coordinating such conceptual operations would explain the way Kit expressed herself in Excerpt 3.

Excerpt 4 (below) continued the discussion around the question of what it means to think that the event in question is unusual. It highlights Luke's thinking with regard to similar issues experienced by Kit, and it illustrates the instructor's effort to scaffold the development of that thinking towards an operational image involving the composition and coordination of the requisite images discussed above.

Excerpt 4 (Lesson 6)

425. Instr: All right, go ahead, Luke.
426. Luke: In a collection of nights it would be unusual if a majority or uhh, it didn't, it would be unusual if most of the time he'd see, or 2 in (inaudible)
427. Instr: Ok, so you, there you're thinking suppose that we've got, we're looking back at the past month.
428. Luke: Right.
429. Instr: We're looking back at 30 nights. And so it would be unusual, and so then you could sort of check them off: "saw at least two people," skip, skip, skip "saw at least two people" (gestures as though moving along days in a calendar and making a check mark), skip, skip. So out of those 30, it would be unusual for him to see uhh two or more people if, if what was true about those 30 nights?
(3-second silence)
430. Luke: That he saw two or more people?
431. Instr: Yeah. No, what—?
(Several students chuckle at the apparent confusion)
432. Luke: I didn't follow your question.
433. Instr: All right. We've got those 30 nights. We go along and we check, we've got, like, a board that's numbered 1 through 30
434. Luke: A calendar!
435. Instr: and we check those nights. Every night that he sees two or more people, he puts a check (motions with hand as though making a check mark).
436. Luke: All right.
437. Instr: All right. So what would it then mean about those 30 nights that it's unusual for him to see at least 2 people?
438. Luke: Most of the time he didn't see two or more people.
439. Instr: Yeah, or most of those nights aren't checked.

In Excerpt 4 Luke began his attempt (line 426) by referring to a "collection of nights," suggesting that he had an image of collecting past data on which to base an assessment of unusualness. But Luke wasn't quite able to coherently articulate what he would look for in those past nights; he appeared to confound "unusualness" with a tenuous description that seemed more consistent with a meaning for "usual." The instructor offered support (lines 427-429) by describing a scenario in terms of a structuring metaphor intended to help Luke imagine what he would keep track of on those past nights. Luke's response (line 430), however, suggests that he was attending only to the event of interest "he saw two or more people" and not coordinating it with an image of the frequency with which the event might occur. Eventually, in line 439, Luke was able to successfully coordinate all the parts; his response suggesting that he understood "unusual," at least momentarily, to mean that on most of the nights Ephram did not see two or more acquaintances at the movie theatre.

Luke's thinking in Excerpt 4 illustrates the slipperiness of the emergent images and the precarious control that students generally had of their composition; they often struggled to piece together component images to create and articulate a coherent operationalization of the event's unusualness as a quantity. In their attempts to imagine

the scenario in a manner that might support quantifying the event's unusualness, their focus of attention would easily and unwittingly shift among the component images, thereby making it very difficult to compose them into an integrated and stable whole. Students were eventually able to provide more or less coherent articulations of unusualness in quantitative terms, but they did so largely under heavy scaffolding by the instructor as typified in Excerpt 4.

To summarize, the discussion excerpts presented in this section illustrate the students' difficulties in thinking of, and describing, an event's unusualness in quantitative terms. Moreover, they indicate that students were not easily oriented toward construing each of the activity situations (e.g., the birthday problem and the movie theatre scenario) in a manner amenable to quantifying an event's unusualness. I hypothesize that a key factor implicated in this difficulty was a lack of explicit cues in the problem situations that might have evoked students' prior knowledge of repeated sampling as a basis for quantifying an event's likelihood, and its potential unusualness. In retrospect, the discussions around the simulation design activities thus serve as an indicator of the limits of the students' prior knowledge; they suggest that a good number of the students had not reflectively abstracted (Piaget, 2001; von Glasersfeld, 1995) the essence of their method of repeated sampling employed in the preceding phase of the experiment to operationalize and generalize unusualness as a statistical quantity. This claim is further supported by students' written responses to a post-experiment test question administered at the end of this phase of the teaching experiment, after eight days of instruction aimed specifically at fostering their ability to conceive a sampling outcome's potential unusualness as a statistical quantity and to describe such in operational terms. Table 1 displays students' written responses to the test question (Figure 4) that queried their thinking with respect to this goal.

Table 1. Student responses to the post-test question

Student	Response *
Nicole	An unusual event would be one the most unlikely to occur. (I.E.- In 3-card poker it is unusual to get a 3 of a kind.) It's that something that occurs that wasn't predicted.
Sue	That mean there is a only few percentage of event occur during the longrun collection of samples.
Kit	When something happens a small % of the time.
Sarah	Of all the samples taken, or items tried, the usual occurance happens the least or close to the least.
Peter	That an event is not likely to happen. If 1000 samples are taken and unusual event will happen about 5% or less of the time.
Tina	It means that., How unusual is it for this to occur? That it does not occur/show up as often.
David	The event unusual means in statistics that there is a lower percentage ["chance"] that something unusual will occur. Like something unusual will probably only occur 10% of the time.
Luke	In statistics the word usually relates to the occurance that something occurs. When someone says that the results are unusual, then they mean that the results don't come out like this on a common occurance.

* All responses are presented verbatim and have not been edited for grammar or punctuation.

What does it mean, in statistics, that an event is “unusual”? (We know that *unlikely*, *unexpected*, and *rare* are synonyms of unusual, so mentioning them will not answer the question. Please explain the meaning, don’t just give synonyms.)

Figure 4. A post-experiment test question posed to students

These responses were analyzed for whether they explicitly mentioned and coordinated two key ideas previously discussed as necessary for quantifying unusualness: 1) an idea of collecting multiple samples or repeating a process multiple times, and 2) an idea of infrequency (or low relative frequency). Three of the responses—Sue’s, Sarah’s, and Peter’s—entailed both ideas, thus suggesting those students’ emergent abilities to quantify unusualness, and their having understandings consistent with that targeted in instruction. Kit’s, Tina’s, and David’s responses referred explicitly to the second idea, but only implicitly to the first. These three responses were coded as *ambiguous* because I could not ascertain how elaborate their images of “percent of the time” or “does not occur as often” were. Finally, Nicole’s and Luke’s responses contained far less evidence of an ability to operationalize “unusualness,” and were thus coded as inconsistent with the targeted understandings. More specifically, Nicole’s response amounted, in places, to a rephrasing using synonyms for “unusual,” whereas Luke’s response was circular in its use of the term. Although both referred to the idea of an event or occurrence being uncommon or not predicted, they did so in a manner that is non-quantitative according to the criteria above.

5.2. THEME 2: CONSTRUING SITUATIONS AS STOCHASTIC EXPERIMENTS

The discussions around the movie theatre scenario (see Figure 2) brought to light other difficulties that students experienced in re-conceiving a given situation as an idealized stochastic experiment. These difficulties can be broadly interpreted as problems of constructing a mathematical model (Doerr & Pratt, 2008). The tasks necessitated construing and re-describing the given situations in terms of idealized assumptions, involving the identification of a suitable population, a sample, and a random sampling process. The activity scenarios did not describe such aspects and relations per se. Instead, students had to learn to construe scenarios in those terms—a process that entailed re-configuring and creatively interpreting the given information. This turned out to be a significant challenge for most students. As illustrated in the following data excerpts, their progress was tentative and tightly embedded within their interactions with the instructor in the classroom discussions.

Constructing assumptions The following illustrative sequence of data excerpts is from the classroom discussions about explicating the assumptions for simulating the movie theatre scenario. Excerpt 5 illustrates that David was overwhelmed by the possibilities for assumptions.

Excerpt 5 (Lesson 6)

428. David: I didn’t get this question ‘cause it, there are so many different things that could happen. Like, what if only half the town goes to see movies? Or uhh what if it’s the same 2 people every night, that he sees? It says he knows 300, but couldn’t, like, the same 2 people go see the movie every night?
429. Nicole: Yeah.
430. Instr: Sure, that’s right. So that’s where you lay—
431. Peter: Good thinking!

432. Instr: you settle all of this in your assumptions. Like, one of the assumptions that you have to make in order to look at this in the abstract, without actually knowing him and the town, is that it's a random process by which the veranda, uhh theatre gets filled every week. (3-second pause) Now, it may not in fact be! But in, that's an assumption that you could make that will let you proceed.
433. David: Oh, ok.
434. Sarah: You also have to assume that he sees everyone that goes to the movie.
435. Instr: Very good! Because if he only sees a small fraction of the people going in, people could be there and he might not see them. (3-second pause) All right. So we're not saying he does, but we're saying in order to proceed we'll make this assumption. Ok, all right. Does that make sense, David?
436. David: Yeah.

Excerpt 5 indicates that David was unable to decide what to assume because he felt lost in a myriad of possible choices. David seemed to view an assumption as a hard fact about the situation, rather than as a working supposition upon which to proceed further. Thus, David's difficulty appeared to be in looking beyond the information given in the situation and reconfiguring it in terms of aspects that are not explicitly given per se, but which are nonetheless necessary to presume. The need to reason hypothetically about a situation as a starting point for designing an investigation of an issue, together with the absence of clear constraints on what could be hypothesized, made the tasks seem too open-ended and ambiguous to some students. The classroom discussions were intended to help students learn to deal with such ambiguity by providing them with opportunities to unpack their implicit assumptions and create new assumptions.

Students' difficulties in deciding what to assume about the given situations were ongoing in these discussions. Decisions were rarely made in a clear-cut manner. Instead they often emerged out of relatively arduous negotiations embedded within messy interactions. The subsequent excerpts illustrate this. Excerpt 6 begins with David struggling to make sense of the underlined part of the central question that was posed as part of the movie theater scenario: "*Is it in fact unusual that at least two people Ephram knows attend the movie he shows, or could people be coming because he is there?*"

Excerpt 6 (Lesson 6)

482. David: Why did you throw in that last part that says "or could people be coming because he is there?" Why did you put that part? That was, that wiggled me out, I didn't know what to do. It's, like, what is that?
483. Instr: Oh! Well—
484. David: It says (reads) "or could people be coming just because he is there?"
485. Nicole: Yeah! That's my point!
486. Instr: Or for some other reason or another—
487. David: Yeah. I was, like, what is that?
488. Instr: Well, if he always saw—
489. Peter: We have to assume that they're not?
490. Instr: if he always saw 30 people that he knows—a tenth of the people that he knows in this town are there every night, then something's going on, right? (2-second pause).
[...]
491. David: Yeah, maybe he's sneaking them in for free.
492. Instr: Perhaps. Something's going on. Would you expect him to see very many people that he knows? If he knows, if there are 30,000 people and it's a random draw to fill the theatre, would you expect him to see very many people that he knows?
493. Nicole: Well, how many movies are there a night?

494. David: He only knows, like, 1% of the town, so it's kind of weird that he'd see people, 2 people every single night.
495. Luke: Yeah.
496. Instr: Yeah, he knows 1% of the town.
497. Instr: Well, we're going to simulate it, we don't know the answer to the question yet!
498. Kit: I think there's one movie per night.
499. Instr: It might be rare.

The question that David and others found so problematic in Excerpt 6 was intended to provoke reflection on the reasonableness of the random attendance assumption, in case the event in question turned out to be statistically unusual. However, this appeared to David and Nicole (lines 482-487) as an isolated question that made little sense and which they could not relate to the greater task. In retrospect, their difficulty is not surprising; since the class had not yet empirically investigated whether the event in question might be unusual, these students could not see the relevance of this statement. The tension that these students experienced drove the instructor to start bringing issues of the underlying assumptions out into the discussion.

Conceiving a population and a sample In the ensuing interaction in Excerpt 6 (line 494), David expressed his gut feeling that the event of seeing two or more acquaintances at the movie theater each night is surprising given the assumption that Ephram knows only 1% of the 30,000 people in the town. Thus, David's intuition touched on the idea of drawing a non-representative sample from the underlying population. Moreover, David's intuition also suggests that he was mindful of the 300 acquaintances as a proportion of the entire population, but that he did not reason similarly about a sample to think of the number of acquaintances as a proportion of the 250 people selected. Had he done so and noted that two or three people out of 250 is close to 1%, he might have been less surprised by such a result. Nicole (in line 493) also began wondering what to assume about the number of movies shown each night, suggesting her effort to conceive the scenario in terms of an unambiguous sampling experiment (perhaps one in which the act of looking into a theater containing 250 people once per night is *as if* one were recording the composition of a 250-person random sample collected from the town's population).

In Excerpt 7 the instructor moved to engage students in choosing values for the Prob Sim parameters in order to simulate the movie theatre scenario. The discussion thus turned to making explicit connections between the scenario and an idealized sampling experiment, guided by the structure of the software interface.

Excerpt 7 (Lesson 6)

504. Instr: Now, I'm going to, I'm gonna do this in a way that uhh the guy who wrote this program suggested (sets up Prob Sim to simulate movie theatre scenario). Put a little tiny dot to represent a person in the town who he doesn't know, and a big dot to represent a person in the town that he does know. How many of these dots are there gonna be in this mixer? (points to small dot in left-most element label on the screen displayed in Figure 5)
505. Luke: 30,000
506. Instr: No
507. Other: 27,000
students
508. Instr: Yeah, 27,000. No, 29,700
509. Peter: 29,700
510. Instr: (enters 29,700 into the "how many" slot under the first element label in Prob Sim; see Figure 5)

511. Peter: Mathematical geniuses! (laughter in background)
 512. Instr: So how many people does—uhh that's because he knows 300 of those 30,000. Right?
 513. Luke: He knows (inaudible) thousand.

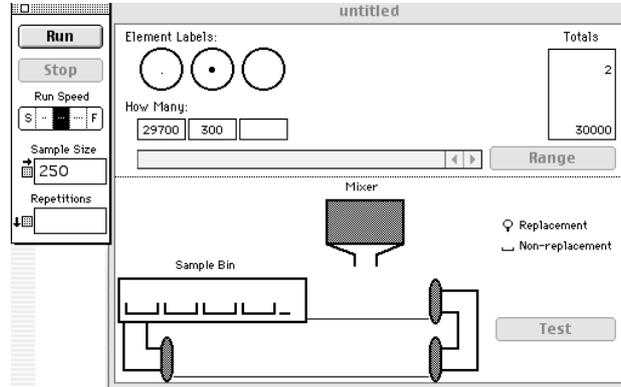


Figure 5. Prob Sim parameter settings for simulating the movie theatre scenario.

- [...]
 514. Instr: Ok, we're gonna take 250 of those people. Right? (assigns this value as sample size in window on screen) Are we taking them with or without replacement?
 515. Lesley: Without.
 516. Peter: With, with!
 517. Luke: No, without replacement.
 518. Lesley: With!
 519. Nicole: No, you can't, it
 520. Luke: You can't—
 521. Nicole: If it's people it has to be—without
 522. Peter: Because they can come back the next night.
 523. Instr: No, no, we're talking about one night.
 524. Kit: Yeah, but not on the same night.
 525. Peter: Oh!
 526. Luke: Repetitions is (inaudible).
 527. Instr: One night. So, is it with replacement or without replacement? (points cursor at "replacement" option in Prob Sim window on screen; see Figure 5)
 528. Nicole: With.
 529. Lesley: With.
 530. Luke: Without replacement.
 531. Kit: Without.
 532. Instr: If a—can a person be in a theatre twice?
 533. Kit: No.
 534. Luke: No.
 535. Lesley: No.
 536. Instr: Ok, so it's without replacement.
 537. Peter: They snuck back in and watched it again.
 538. Nicole: Wait.
 539. Kit: Not at the same time you can't.
 540. David: You got it on without replacement.
 541. Nicole: It's one movie one night.
 542. Instr: Yeah.
 543. Lesley: I don't understand.
 544. Instr: Or, it's just one night. We don't know how many movies, but—

545. Nicole: Well then it's a difference! That's what I asked you
 546. Instr: Ok, then let's say one movie one night. That's a good assumption.

Excerpt 7 began with the instructor using a small dot and a large dot in the software's element labels to represent the population items *non-acquaintance* and *acquaintance*, respectively (see Figure 5). Luke then proposed that the Mixer—Prob Sim's representation of a population—should contain 30,000 small dots (line 505), suggesting that he had not yet conceived of the population as comprised of those two distinct classes of items. The instructor and other students chimed in with a different answer (lines 506-509), and after a mental computation error was resolved, the instructor explained his choice of the number of small dots in terms of an appropriate population proportion.

The discussion then turned to assigning the sampling parameters. The instructor proposed that sample size should be 250 (line 512), and the issue then became whether to sample with or without replacement. Here, direct evidence of students' thinking about the issue is limited. However, students clearly had different different opinions about this, and I hypothesize that the source of those differences was rooted in their different (implicit) assumptions about how many movies are shown each night. Peter's utterances in lines 522 and 537, together with the surrounding student contributions, provide insights into his thinking. In line 522, Peter appeared to assume that people could return to the movie theater on different nights, suggesting that he had not yet structured the situation in terms of what might occur on an individual night as a distinct unit. Instead, he seemed to be considering events that could occur across several nights. Peter's utterance in line 537 indicates that even when restricting himself to considering an individual night, he was thinking of contingencies that suggest he was unclear as to *what* constituted a sample in the scenario. Peter's difficulty can be interpreted as one of making an idealized assumption about the situation. His comments presumably reflect the challenge he experienced in construing the situation as an unambiguous sampling experiment in which a sample consisted of a 250-person audience that attended a single movie on one night.

Although the best available evidence is of Peter's thinking, Excerpt 7 illustrates that the group was generally indecisive about whether to sample with or without replacement. Students who flitted from one sampling option to the other (i.e., Lesley and Nicole) were evidently unsettled about what to take as a sample in the scenario; their assumptions were still formative and highly unstable. Eventually, the instructor and Nicole settled on the assumption that a sample should correspond to a single movie on a single night (lines 541-546)—a simplification consistent with sampling without replacement.

To summarize, the data excerpts discussed in this section provide important glimpses into how construing a contextual situation as a stochastic experiment can be a highly non-trivial activity for students. Even under conditions of heavy instructional guidance, involving a supportive classroom environment in which it was normative to dissect ideas, and the use of software intended to structure and constrain the activity in productive ways, students experienced considerable difficulties. Salient among these was difficulty in making discriminating assumptions about the situations to be modeled. The connections between a given situation *as it is described* and its construal as a stochastic experiment were generally not transparent for students.

6. DISCUSSION AND CONCLUSION

The findings of this study center around two themes that speak to the challenges of i) conceiving of expectation as a statistical quantity, and ii) conceiving of situations as stochastic experiments. Regarding the first theme, on a broad level the findings

underscore simply how difficult it can be for students to move from having an intuitive sense of an event's expectation to quantifying that sense. On a deeper level, interpreting students' difficulties through the frame of quantitative reasoning (Thompson, 1994) reveals some of the finer intricacies of those difficulties that move us toward a level of explanation for them. At the core of students' difficulties in quantifying an event's likelihood lay persistent and robust challenges in conceiving of the underlying situations in a manner that would support their thinking of the expectation of an event embedded within them as quantifiable. More specifically, the key conceptual operations ostensibly entailed in doing so that posed significant challenges for students are: imagining a given situation in terms of an idealized sampling experiment that could be repeated under the same conditions; generating an image of the anticipated accumulation of those repetitions, and generating an image of the anticipated accumulation of the number of times an event of interest is observed; and composing these images so as to quantify expectation of the event of interest as its anticipated relative frequency. With regard to the last three operations, the difficulties exhibited by students Kit and Luke (highlighted in data Excerpts 3 and 4) provide a paradigmatic illustration of how students' efforts to generate and compose such images were fraught with an instability, wherein their emergent images tended to unwittingly "dissolve" into one another. Additionally, as indicated by the results of the post-test question that queried students' abilities to operationalize "unusual event" (see Table 1), a majority of the students continued to struggle to develop these operations even after having participated in eight instructional sessions dedicated to fostering such development. The lack of control that students exhibited in generating and coordinating the requisite images is indicative of their difficulties in conceiving the repeatable sampling process and the event of interest as identifiable and distinct conceptual entities. The absence of a clear mental image of the two entities would consequently make it difficult to imagine their co-accumulation and the event's expected relative frequency.

The first of the above-mentioned conceptual operations (imagining a given situation in terms of an idealized and repeatable stochastic experiment) relates directly to the second theme of the study's findings. The difficulties documented here indicate that it was not straightforward for students to construe a contextual situation in terms of a population or a sample drawn from it, in a manner that could enable them to reconceive the situation as a repeatable sampling experiment. At the core of students' difficulties were ambiguities regarding what to conceive as the "repeatable entity," seemingly driven in part by their uncertainties concerning what idealized assumptions to make about the underlying situation and how to map such assumptions to methods of sampling (with or without replacement). As illustrated in the discussion excerpts, collectively students moved toward eventually constructing such assumptions and mappings through an arduous process of negotiation and only under tight scaffolding by the instructor.

The conceptual operations discussed above can be seen, more broadly, as key entailments of conceptualizing a *trial* of an experiment—"an instantiation of an experiment that yields a public outcome" (Horvath & Lehrer, 1998, p. 122) so that two instantiations of it may be compared, assuming they are essentially identical with respect to some key structures. The issue of what constitutes a trial of an experiment for learners was not an explicit *problématique* of the studies discussed at the beginning of this report. Instead, the idea of a trial of appears to have been taken as a given in most of those studies, as reflected by their use of particular tasks and computer environments that were evidently not intended to have students confront and problematize the idea. The evidence presented and discussed in this study, however, arguably provides good reason not to take the idea of a trial as a given in stochastics education, but rather as one that learners must

construct and whose construction can entail considerable conceptual machinery and instructional challenges. With regard to this last point, the findings of this study point to a need for research that focuses on an aspect not directly under its purview, but nevertheless informed by its findings—namely, instructional design that takes the development of a stochastic conception of sampling as an important problem to address. The importance of such research is further underscored by the increasing prevalence of new statistics curricula in which the use of stochastic simulations is prominent (Garfield et al., 2012; Rossman & Chance, 2012).

ACKNOWLEDGEMENTS

Research reported in this paper was supported by the National Science Foundation (Grant No. REC-981187). Any conclusions or opinions stated here are the author's and do not necessarily reflect official positions of NSF. The author wishes to acknowledge the contribution of Patrick Thompson, principal investigator of the project from which the data for this study were drawn.

REFERENCES

- Biehler, R., & Prömmel, A. (2010). Developing students' computer supported simulation and modelling competencies by means of carefully designed working environments. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php [© 2010 ISI/IASE]
[Online: http://iase-web.org/documents/papers/icots8/ICOTS8_8D3_BIEHLER.pdf]
- Chance, B., & McGaughey, K. (2014). Impact of simulation/randomization-based curriculum on student understanding of p-values and confidence intervals. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics, Flagstaff, Arizona, USA, July, 2014*. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://icots.info/9/proceedings/pdfs/ICOTS9_6B1_CHANCE.pdf]
- Cortina, J. L., Saldanha, L. A., & Thompson, P. W. (1999). Multiplicative conceptions of arithmetic mean. In F. Hitt & M. Santos (Eds.), *Proceedings of the Twenty First Annual Meeting: North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 466-472). Cuernavaca, Mexico: ERIC Clearinghouse for Science, Mathematics, and Environmental Education, Columbus, OH.
- delMas, R. C., Garfield, J., and Chance, B. L. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education*, 7(3),
[Online: <https://ww2.amstat.org/publications/jse/secure/v7n3/delmas.cfm>]
- Doerr, H. M., & Pratt, D. (2008). Mathematics and mathematical modeling. In M. K. Heid & G. W. Blume (Eds.), *Research on technology and the teaching and learning of mathematics: Research syntheses* (Vol.1, pp. 259-285). Charlotte, NC: Information Age Publishing.

- Drier, H. S. (2000). Children's meaning-making activity with dynamic multiple representations in a probability microworld. In M. Fernandez (Ed.), *Proceedings of Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 691–696). Columbus, OH: ERIC Clearinghouse of Science, Mathematics, and Environmental Education.
- Finzer, W. (2012). *Fathom dynamic statistics* (v.2.1). Computer program. Key Curriculum Press.
- Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM-The International Journal on Mathematics Education*, 44(7), 883-889.
- Harel, G., & Koichu, B. (2010). An operational definition of learning. *Journal of Mathematical Behavior*, 29, 115-124.
- 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, NJ: Erlbaum.
- Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: Responding to classroom realities. In F. Lester, (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909-955). Reston, VA: NCTM.
- Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, 11(2), 143-157.
- Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6(1), 59–98.
- Konold, C. (2002). *Prob Sim user guide*. Amherst, MA: Author.
- Konold, C., Harradine, A., & S. Kazak (2007). Understanding distributions by modeling them. *International Journal of Computers in Mathematics Education*, 12(3), 217-230.
- Konold, C., & Miller, C. (1996). *Prob Sim*. Computer Program. Amherst, MA.
- Konold, C., & Miller, C. D. (2011). *TinkerPlots™* (v.2.0). Computer program. Emeryville, CA: Key Curriculum Press.
- Kuhn, M., Hoppe, U., Lingnau, A., & Wichmann, A. (2006). Computational modelling and simulation fostering new approaches in learning probability. *Innovations in Education and Teaching International*, 43(2), 183–194.
- Liu, Y., & Thompson, P. W. (2007). Teachers' understandings of probability. *Cognition and Instruction*, 25(2), 113-160.
- Maxara, C., & Biehler, R. (2006). Students' probabilistic simulation and modeling competence after a computer-intensive elementary course in statistics and probability. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. Salvador, Brazil. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://iase-web.org/documents/papers/icots7/7C1_MAXA.pdf]
- Maxara, C., & Biehler, R. (2007). Constructing stochastic simulations with a computer tool: Students' competencies and difficulties. In D. Pitta-Pantazi & G. Philippou (Eds.), *Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education* (pp. 762-771). Larnaca, Cyprus. University of Cyprus, Department of Education.
[Online: <http://www.mathematik.uni-dortmund.de/~erme/CERME5b/WG5.pdf>]
- Mills, J. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education*, 10 (1).
[Online: <http://www.amstat.org/publications/jse/v10n1/mills.html>]
- National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

- Noll, J., Gebresenbet, M., & Glover, E. D. (2016). A modeling and simulation approach to informal inference: Successes and challenges. In D. Ben-Zvi & K. Makar (Eds.), *The teaching and learning of statistics: International perspectives* (pp.139-150). New York: Springer.
- Piaget, J. (2001). *Studies in reflecting abstraction* (R. L. Campbell, Ed., Trans.). New York: Psychology Press.
- Powell, A. B., Francisco, J. M., & Maher, C. A. (2003). An analytical model for studying the development of learners' mathematical ideas and reasoning using videotape data. *Journal of Mathematical Behavior*, 22(4), 405-435.
- Pratt, D. (2000). Making sense of the total of two dice. *Journal for Research in Mathematics Education*, 31(5), 602-625.
- Rossman, A. J., & Chance, B. (2012). *Workshop statistics: Discovery with data* (4th ed). Wiley & Sons.
- Saldanha, L., & Thompson, P. (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.
[Online: <http://www.iejme.com/makale/336>]
- Saldanha, L. A., & Thompson, P. W. (2014). Conceptual issues in understanding the inner logic of statistical inference: Insights from two teaching experiments. *Journal of Mathematical Behavior*, 35, 1-30.
- Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical implications*. Mahwah, NJ: Lawrence Erlbaum.
- Shaughnessy, J. M. (1992). Research on probability and statistics: Reflections and directions. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). Reston, VA: NCTM.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. Lester, (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957-1009). Reston, VA: NCTM.
- Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 267-306). Mahwah, NJ: Lawrence Erlbaum Associates.
- Stohl, H. (2002). *Probability Explorer*. Software package (v.2.0).
- Stohl, H., & Tarr, J. E. (2002). Developing notions of inference using probability simulation tools. *Journal of Mathematical Behavior*, 21, 319-337.
- Thompson, P. W. (1994). The development of the concept of speed and its relationship to concepts of rate. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 179-234). Albany, NY: SUNY Press.
- Tintle, N. L., Rogers, A., Chance, B., Cobb, G., Rossman, A., Roy, S., Swanson, T., & Vanderstoep, J. (2014). Quantitative evidence for the use of simulation and randomization in the introductory statistics course. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics, Flagstaff, Arizona, USA, July, 2014*. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://icots.info/9/proceedings/pdfs/ICOTS9_8A3_TINTLE.pdf]
- von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning*. London: Falmer Press.
- von Mises, R. (1957). *Probability, statistics, and truth*. London: Allen & Unwin.

- Watson, J. M., & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31(1), 44-70.
- Well, A. D., Pollatsek, A., & Boyce, S. J. (1990). Understanding the effects of sample size on the variability of the mean. *Journal of Organizational Behavior and Human Decision Processes*, 47(2), 289-312.
- Zieffler, A., delMas, R., Garfield, J., & Brown, E. (2014). The symbiotic, mutualistic relationship between modeling and simulation in developing students' statistical reasoning about inference and uncertainty. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics, Flagstaff, Arizona, USA, July, 2014*. Voorburg, The Netherlands: International Statistical Institute.
- [Online: http://icots.info/9/proceedings/pdfs/ICOTS9_8B1_ZIEFFLER.pdf]

LUIS SALDANHA
Département de mathématiques
Université du Québec à Montréal
C.P. 8888, succ. Centre-ville
Montréal (Québec)
H3C 3P8 CANADA