**Title**
Reasoning about Probabilistic Phenomena: Lessons Learned and Applied in Software Design



**Journal**
Technology Innovations in Statistics Education, 3(2)

**ISSN**
1933-4214

**Authors**
Lee, Hollylynne S
Lee, J. Todd

**Publication Date**
2009-12-23

**DOI**
10.5070/T532000037

Peer reviewed

# Reasoning about Probabilistic Phenomena: Lessons Learned and Applied in Software Design

## 1. TECHNOLOGY ADVANCES IN PROBABILITY TEACHING AND LEARNING

The 21st century is a complex, data-driven technological world. Public literacy and preparation for the workforce demand that students be able to make data-based decisions, risk analysis, inferences, and predictions. All of these statistical pursuits have probabilistic phenomena at their heart. By probabilistic phenomena we mean situations where data is produced via a random experiment repeatedly conducted on a population or space of possibilities. Probability and statistics is now a common curricular strand in K-12 school curricula and there is a rapid increase in students taking the Advanced Placement Statistics exam in high school (e.g., College Board, 2007; Franklin et al, 2005; National Council of Teachers of Mathematics [NCTM], 2000). With access to more advanced technologies, teachers are encouraged to use tools for graphical and numerical data analysis and use an empirical introduction to probability through computer simulations (Batanero, Henry, & Parzysz, 2005; Biehler, 1991; Ben-Zvi, 2000; College Board, 2006; NCTM, 2000; Parzysz, 2003). There is a growing body of literature on students' understanding of probability and statistics concepts (e.g., Jones, 2005; Jones, Langrall, & Mooney, 2007; Shaughnessy, 1992, 2003, 2007), and general agreement that research has been lacking sufficient study of how learners make connections between empirical data generated by repeated random experiments and a theoretical model of probability (e.g., Jones, 2005; Jones et al, 2007; Parzysz, 2003).

In describing the various ways technology tools can support students' learning in statistics and probability, Chance, Ben-Zvi, Garfield, and Medina (2007) promote the use of simulations as pedagogical tools. In particular, they emphasize the role of technology in allowing students to study random processes and to observe what happens when a process is repeated a large number of times. The pseudorandom number generators in technology tools use a function dependent on a defined distribution as the basis for its input to generate subsequent "random" outputs. Thus, in a technology environment, students can model probabilistic situations based on assumptions about a theoretical distribution, simulate an experiment to generate a large amount of data, and manipulate and represent the data in various ways that would be nearly impossible to do within the time constraints of school curriculum and instruction. In this regard, technology offers a rich medium for designing tools and studying students' reasoning about probabilistic phenomena.

As Biehler (1991) noted almost 20 years ago, "It would be valuable to have more experiences with software where students can design random devices on the screen" (p. 189). Since that time, many researchers have designed probability software based on different perspectives and with sometimes different goals for students' learning (e.g., *ProbSim,* Konold, 1992; *Chance World*, Jiang 1992; *Probability Inquiry Environment*, Vahey, 1998; *ChanceMaker*, Pratt, 1998; *Probability Explorer*, Drier, 2000a; *ProbLab,* Abrahamson &Wilensky, 2002; *Model Chance*, Konold & Kazak, 2008). Jones et al (2007), in their review of recent advances on the study of computer software in probability instruction, state:

> The most promising research involves the use of microworlds and their influence of changing and expanding students' probabilistic thinking (Konold, 1995a; Paparistodemou et al., 2002; Pratt, 1998; Pratt & Noss, 2002; Stohl, 1999-2002) [*citations in original*]. In particular, the work of Pratt and Stohl shows that microworlds can be designed to forge vital links between experimental and theoretical probability and to build cognitive mechanisms that enable children to challenge their own intuitions. (p. 946)

Several studies have documented how students are able to make connections between distributions of data from a simulation and the theoretical distribution described in the model with particular attention to the effect of the number of trials (e.g., Abrahamson & Wilensky, 2007; Drier, 2000a; Konold, et al., 2007; Konold & Kazak, 2008; Pratt 2000; Stohl & Tarr, 2002b). Collectively, these studies suggest that these tools give students control over designing experiments, conducting as many trials as they desire, viewing graphical representations of results, and can help students develop understandings of the relationship between theoretical models and empirical data.

In this paper, we have several goals. First, we provide a brief description of the key features of a technology designed to allow young students opportunities to explore probabilistic situations and then explain several research observations made in multiple investigations of student explorations with this probability micro-world software package. These observations are used to posit a model for a way of reasoning about probabilistic phenomena that focuses on the bidirectional interplay between theoretical distribution and empirical data. The lessons learned from students' work on probability tasks and the bidirectional model have implications for instructions that we discuss. The paper concludes with a discussion of a next generation innovation for representing the theoretical distribution in the software and future research directions. This new distribution tool is designed to facilitate the transition from intuitive equiprobable discrete probability distributions to the far more complicated density representations that lie at the heart of statistical investigations.

## 2. DESIGN OF PROBABILITY EXPLORER

Beginning in 1998, H. S. Lee began work in developing the *Probability Explorer* software tool as part of her doctoral dissertation work. The current version of *Probability Explorer* [*PE*] (Stohl, 2002, v.2.01) was a result of prior iterative research and design studies (Drier, 2000a, 2000b, 2000c, 2001; Stohl & Tarr, 2002a, 2002b). *PE* is designed to allow students to explore the numerical representations of an underlying probability distribution, as well as the numerical and graphical representations of the distribution of results from repeated trials (Drier, 2000a, 2001). A probability distribution is currently represented in *PE* with a finite set of outcomes, each of which has an integer "weight". Various numerical representations of this distribution can be accessed through a "Weight Tool" that allows students to examine and create distributions that build from part-to-part and part-to-whole reasoning (see Figure 2). Traditional scenarios of fair coins, dice, and bag of marbles are easily represented; bias coins and dice and several real world scenarios (such as the weather) can also be represented. To conduct a simulation, students decide the possible outcomes for a simple experiment, how many of these to combine into a compound experiment (1, 2, or 3), the number of trials to conduct, how to arrange data, and which graphical or numerical representations to view for analysis (see Figure 1). In addition, all data representations update dynamically *after each trial* to facilitate students analyzing data during a simulation, rather than only viewing representations of data in an aggregate static form (Drier, 2000b). Collectively, these tools can help students grapple with key ideas (sample size, independence, and variability) that are critical in understanding a relationship between empirical data and theoretical distributions (Stohl & Tarr, 2002a, 2002b).

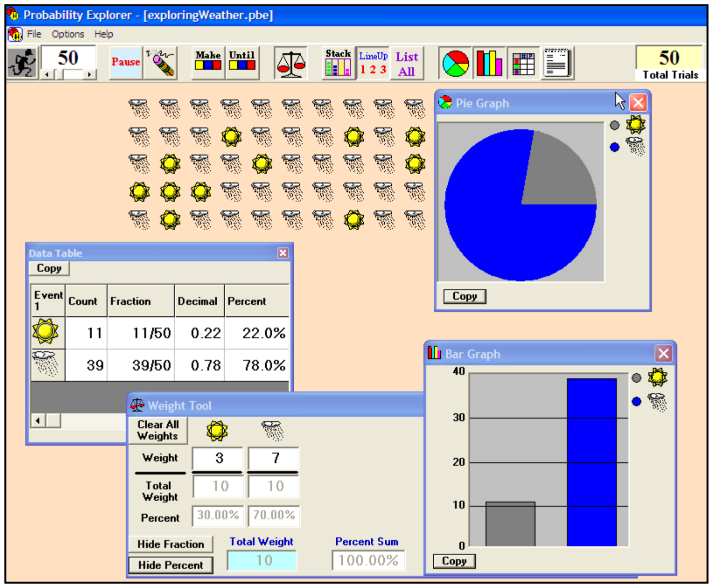

**Figure 1:** Screenshot of sample of 50 trials

## 2.1 Modeling Probability as Perceived in the Physical World

Typical activities in probability involve the use of devices such as coins, dice, marbles, and spinners. It is not our intention to replace physical experiences with digital simulations. Without prior use of such physical devices, children may not fully comprehend the randomness of the computer simulation or make meaningful connections between the two-dimensional icons and their three-dimensional counterparts. In fact, Shaughnessy (1992) suggested that "it is important for us to continue developing connections between concrete simulations and computer simulations in our teaching and investigating the effects of the transition between the two in our research" (p. 485).

*PE* is designed to make the transition between concrete materials and computer simulations as seamless as possible. For example, clicking on the "Run" button in order to simulate a trial represents a conscience action by a student. This action is similar to the purposeful act of rolling a die or flipping a coin and anticipating the trial result. Once a random result appears on the screen, a student can act upon the object to move, sort, organize, or stack or line up in a playful and potentially meaningful manner. Actions on these objects can help the child instantiate the experimental results and build simple (piles) or complex (venn-like sorted groups) re-presentations of the data. The ability to have moveable iconic representations of randomly generated data substantially extends the capabilities of experimentation with physical devices where results are usually only listed or tallied. In this regard, the microworld not only connects with the physical world, but extends the potential actions available in this new probability "world."

## 2.2 Designing Experiments

There are several actions needed for defining the type of experiment to simulate in *PE*. There are two "preset" options for students to run simulations with flipping a coin or rolling a regular six-sided die. In addition, students can also design their own experiments, either by designing a bag of marbles with up to six different colors, or by choosing from approximately fifty icons that will represent the possible outcomes (maximum of eight) of a single random experiment (Figure 2). In all cases, the user must also decide whether to conduct a simple experiment, or to create a compound experiment with two or three of the defined simple experiments. The initial assignment of the theoretical probability distribution is equiprobable such that each outcome has $1/n$ chance for $1 \leq n \leq 8$. The actions of deciding how many outcomes in a simple experiment and how many of these experiments should be compounded together provides students with a moment of reflection to think about how they are using the tools in the computer environment to build a model for a chance situation.

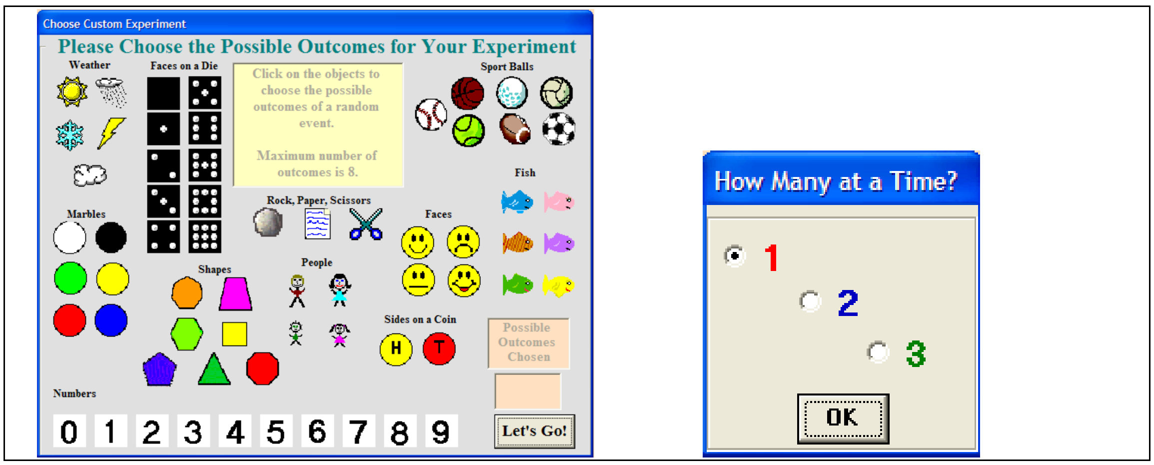


**Figure 2:** Defining the type of random experiment

## 2.3 Modeling a Theoretical Probability Distribution

Real world physical devices such as coins and dice cannot be easily and accurately altered to affect the probability of an outcome occurring. Thus, many students who use such physical devices to model probability situations only experience perceived theoretically equally probable outcomes and may inappropriately apply the equiprobable heuristic to real world situations that are not equiprobable (e.g., it is highly unlikely that two sports teams are truly equally likely to win a game).

As an experiment is being defined, the probability distribution is stored, and can be altered, through a Weight Tool (see Figure 1). The metaphor of "weight" was used in earlier and current designs to help students understand the process of assigning probabilities to an outcome. "Heavier" outcomes are more likely to occur, while "lighter" outcomes are less likely to occur. Weight is measured in units of whole numbers. To facilitate the instantiation of the "weighting" process, students can click on an object in the Weight Tool to increase its weight. Each click corresponds to an increase of one in the weight. By default, students view the distribution of weights as a count. This view of the distribution can facilitate thinking about the part-to-part relationship between the outcomes, a common way that young children may initially think about probability situations (Jones *et al*, 1997, 1999a). This level of thinking is also aligned with children's early fractional thinking when they only consider the "parts" of a fraction (numerator) rather than the "part" in relationship to its "whole" (denominator). A part-to-part display is also similar to the concept of odds and can be useful for distinguishing between the odds and probability of an event. Because theoretical probabilities rely on both the "part" and the "whole," students also have the ability to view the distribution of weights as a fraction and percent. Additionally, children can give all the weight to one outcome and explore the scenario where there are several describable outcomes, but only one of them is possible.

# 3. EXPLORING DISTRIBUTIONS IN *PROBABILITY EXPLORER*

In several prior and on-going studies with students ages 8-14 (e.g., Drier, 2000a, 2000b; Lee, 2005; Lee & Lee, 2009; Lee, Angotti, & Tarr, under review; Stohl & Tarr 2002b; Tarr, Lee & Rider, 2006; Weber, Maher, Powell, & Lee, 2008), it has been observed that many students use similar approaches to interpret the probability distribution (e.g., as shown in the Weight Tool), use similar arguments for choosing the number of times to repeat a random experiment, and often make similar observations regarding the distribution of empirical results. In each of these studies, students were working in small groups (2-3 students per computer) with *Probability Explorer*.

Students' interactions with the software and each other in the primary studies were videotaped and analyzed for critical events (Powell, Francisco, & Maher, 2003). These critical events were then examined more carefully and an open coding was used to develop categories related to students' interpretation and interactions with theoretical probability distributions and empirical distributions. Each of the critical event episodes were then grouped according to category and examined in depth. To illustrate some of the key lessons learned within each of these categories, we highlight several examples of students' playful goal-directed activities, their use of capabilities in *PE*, and their subsequent observations. Each example was chosen to represent similar activities observed across studies.

## 3.1 What is the Role of a Probability Distribution?

In the current design of *PE* (v.2.01), most tasks in which students initially design an experiment involve an equiprobable distribution (e.g., fair coin, fair die, choosing up to 8 possible outcomes), in which the Weight Tool defaults to assigning Weights of 1 to each outcome. When students are asked to interpret the weights, two common interpretations are prevalent:

1) Students often imagine a hypothetical experiment where the sample size is equal to the total weight and they explain that the empirical distribution should be equal, or almost equal, to the assigned weights. For example, in Drier (2000b), Carmella (age 9) designed an experiment with two equiprobable outcomes, the sun and the rain.

Carmella: It means that if you were to press this [points to the "run" button] twice, then one of them would be the sun and one of them would be the rain, most likely.

Teacher: Most likely. Okay and why is that most likely?

Carmella: Because the weight is one and one. And then the total weight would be two. And one is divided, and two is divided into one. And that's most likely because there is no guarantee.

2) Students often describe the weights in terms of an imaginary box or bucket filled with the number of each items equal to each assigned weight. For example, in Drier (2000a), when Jasmine (age 9) designed a weather situation where it would be twice as likely to be sunny than to rain, after a lot of struggles, she assigned 24 to "lightening" and 48 to "sun."

Teacher: 48 and 24. So what do these numbers mean here? 48 over 72 [displayed as a fraction in Weight Tool]?

Jasmine: Forty-eight over 72 …Oh, there are 72 suns and lightning bolts put in the box. Forty-eight of them are suns. Twenty-four of them are lightning bolts. And children put in that many because they think out of 72 days there are going to be 48 and 24, there are going to be 48 sunny days and 24 thundering days.

Of course Jasmine also used the first strategy to apply the distribution to an imagined empirical situation. Neither of these strategies is surprising, as it is common for students to describe the probability of an event in both ways, the first representing a typical empirical probability interpretation, and the second a classical counting approach to computing probability. However, of interest to us is how students actually collect empirical data and any connection they may make to the probability distribution.

When conducting a simulation where a total weight is known from the Weight Tool or suggested from the task context (e.g. knowing there are *n* marbles in a bag but not the exact distribution of colors of marbles), the students gravitate towards using the total weight (or *n*) for an initial sample size (e.g., Lee, 2005). Then, when these same students are conducting simulations with a hidden weight tool and the experiment suggests no integer weights or a total weight (e.g., many fish in a pond), students often are faced with a dilemma and do not know what sample size to choose for their experimentation (Stohl & Tarr, 2002b). At this point a teacher or peer typically has to tell them they need to choose a sample size, or one is suggested.

Students readily accept a strong tie between the Weight Tool and trial results. This is a strong start in forming intuitions about the connections between the theoretical distribution and empirical data. Students are not surprised by some variation between samples and their expectations based on the Weight Tool. However, a big question is what tolerance do they have for this variation? How do they form intuitions about this variability? In fact, in many episodes we observed, students observing surprising variations provoked playful exploration.

> *A key lesson learned is that the integer values in the Weight Tool are highly suggestive to students, both in interpretation and sample size choice. And it may be these interpretations that drive student expectations for a rather close match between the empirical data and the distribution of weights.*

## 3.2 Search for an Exact Empirical Distribution

When students have enough queues from the distribution to employ exact, small sample size selections (like choosing a sample size of 4 when drawing from a bag of 4 marbles), they discover that often they will be rewarded with exact matches to the theoretical distribution (Drier 2000a). However, they also discover fairly quickly that such small samples can yield data that varies greatly from their expectations (Lee, Rider, & Tarr, 2006). In multiple episodes, students have demonstrated strategies of lowering the number of trials in a sample in order to have a higher probability of getting a particular distribution, typically one that is a mirror image of the probability distribution, particularly with experiments of 2-3 outcomes in the sample space (e.g., Lee, 2005). At first glance, giving students the freedom to choose the number of trials seems likely to reinforce their use of representativeness heuristics (Tversky & Kahneman, 1971, 1982). But their reasoning is correct. For example, drawing with replacement from a bag containing 50% black and 50% white marbles, it is more likely to obtain a half white and half black marble distribution from a sample of 4 trials ($p=0.375$) than it is with a sample of 10 trials ($p\approx0.246$).

> *A key lesson learned is that the goal of finding an exact empirical distribution, and the strategy of lowering the number of trials in a sample to increase the probability of obtaining that distribution, is evoked when students know the underlying probability distribution, particularly if it is an equiprobable one.*

## 3.3 Noticing Variability with Small Samples

In many of the same instances where students have a goal to find an exact empirical distribution, they also notice they are more likely to obtain distributions that are very different from the theoretical probability distribution with smaller sample sizes (e.g., Lee, 2005). They begin to attend to variability and the differences in empirical distributions across repeated samples. Students often lower the number of trials to meet a different type of student directed goal. As reported by Lee (2005), two seventh grade students (age 12-13), Jerel and Chris, were using *PE* to draw a marble from a bag with replacement, in which they knew the bag contained 4 blue, 3 green and 3 red marbles. The students initially ran several samples of 10 trials. However, while they were running samples of size 10, they had created a goal of seeing which color “won” with each of them cheering for a color (Jerel claimed blue and Chris claimed red). In these samples of 10, some contained more red, some more blue, and one had a lot of green.

The students decided to run 100 trials and “cheer” for their respective colors. Although the frequency of each color initially changed as well as their relative position in the “race” they were envisioning, there were many more blue marbles after about 75 trials. At this point Chris noted that blue would win and stopped cheering for his color—of course Jerel still cheered on blue and gloated about his win to Chris. When the teacher asked the class if blue was always the marble picked the most, Chris replied “No, because at the beginning we started getting more red and now we are getting more blue.” He noticed a long-term trend in the results from the simulation. All of this is being said to provide a context for his next goal-directed activity. Chris reduced the number of trials to 10, ran the simulation and obtained 5 red, 2 green, and 3 blue—and of course gloated to Jerel. Chris appears to have noticed the power of small numbers for meeting a specific goal—winning the race with red. Chris’ strategy is in accord with the notion that it is with small-sized samples where we are more likely to see outcomes with lower theoretical probabilities occurring more frequently. Whereas, the law of large numbers tells us that the likelihood of this happening limits to zero as the trials size increases. In cases where students only know how many outcomes there are in an experiments’ sample space, some students will conduct several samples of $n$=N trials, notice variability across samples, and then do some type of “averaging” to estimate a distribution (not always formal, and often more aligned with modal clumping). In many cases, this strategy will work and allows for a somewhat reliable estimate.

> *A key lesson learned is that allowing students to engage in playful open exploration that includes freedom of choice of numbers of trials, the ability to*

*quickly do repeated samples, and visualization of data in graphical form can lead them to develop intuitions of the role of small sample sizes in variability that occurs within and across samples.*

## 3.4 Noticing Variability with Large Samples

As shown in the example of Jerel and Chris, they initiated the action of running a larger number of trials during their game-like activity. In several episodes, we have observed students setting goals and creating competitions or games in probability contexts that were not initially posed as such. This speaks to the playful nature of their engagement with the software and the tasks. When self-selecting to run a large number of trials, students often notice something about the distribution of their empirical data, particularly from their interpretations of the graphical representations (pie and bar graph). Consider Amanda's (age 9) work when she had a small number of trials currently displayed in a pie graph from a simulated fair coin toss. She made a goal of trying to create a pie graph that was all blue (i.e., only one outcome) like she had seen in prior results with a 4-5 trials (Drier, 2000a). Even though the data showed some heads and tails (part blue and part gray shown in pie graph) she began to repeatedly press the Run button to add on 10 trials at a time to sample. The action of adding more trials to the data set became a pleasurable experience that led to observations by her and others (all 9-10 years old) about the empirical distribution.

Amanda: Well, it's [size of sectors in the pie graph] staying in the same place pretty much.
Teacher: Why do you think it's staying in the same place?
Amanda: Because…
Carmella: Because she's running it so many times, it's like evening out.
Teacher: Really? Why is it evening out?
Carmella: Because it's so many of them and …
Jasmine: Look how much you've done it [about 1000 trials now]. It's still going.
Teacher: So, Amanda do you think you're ever going to get all blues (heads) or all grays (tails)?
Amanda: No.
Teacher: Why not? Why couldn't we have a pie graph be all blue or all gray?
Carmella: Because it evens out with how many you do.
Teacher: Why did you say that?
Carmella: Because the more you do, the more the chance to even out…with more coins one is still going to be a little bit ahead of the other mostly, but it's unlikely that one [points to sector in pie] will rise a lot above the other. There's so many it can't do it [pause] it's like it evens out.

Amanda's goal, the visualization with the pie graph, and subsequent questioning from the teacher appeared to prompt Carmella's observation and statement, which appears to be an early verbalization of the law of large numbers from a child's perspective. Consider the similarity in other students' descriptions:

Donovan (age 12, fair coin): "When you program it to do 500 and run it, after awhile you can see the pie graph pretty much staying in the center, so it's [pie graph sectors] always even almost and not going back and forth."

Dean (age 8, fair coin after 500 trials): Dean pivoted his forearm to mimic the changes in the graph with large swings [representing the wide variability with a small number of trials] decreasing toward tiny up-and-down motions [to represent the approach to the 50%-50% ratio] with a large number of trials. (Drier, 2000c)

Jon (age 8, fair dice): “The higher you go the more it looks like what you are supposed to get.” (Drier, 2000c)

Many students’ first verbalizations of this phenomenon have typically occurred in equiprobable situations with two outcomes such as fair coin tosses or a bag of marbles with 2 colors, although they are typically able to generalize this phenomenon to non-equiprobable distributions. It seems that equiprobable situations may be helpful in promoting a sense of “evenness” in results from large samples. However, the dynamic visualization of the reduced fluctuation in results, particularly with a pie graph, speaks to students viewing this phenomenon as not only an end result after a large number of trials, but a process that occurs during experimentation (Drier, 2000b). The dynamic pie graph also seems to be the more salient graphical representation than the bar graph in promoting the observations related to a stabilization process. We hypothesize this is related to the proportional nature of the pie sectors. The bar graphs always appear to change due to the rescaling as a sample size increases and the constant height comparisons across bars for the discrete outcomes. Any representation that display results from experiments during the simulation process can focus students’ attention on variability by “observing the fluctuation of samples… and observing the stabilization” (Parzysz, 2003, p. 1) within a large run of trials. Using simulation software affords students an opportunity to observe the dynamic accumulation of data in numerical and graphical forms while data is being generated. As illustrated in the examples of students’ work, this visualization of the data has shown to be a powerful motivator for students noticing variability in short and long term behavior of random experiments (Stohl & Tarr, 2002b) and using their observations as evidence to support their claims about an unknown probability distribution (Weber, Maher, Powell, & Lee, 2008).

*A key lesson learned is that allowing students to engage in playful open exploration with contexts that include equiprobable outcomes, freedom of choice of numbers of trials, the ability to quickly do repeated samples, and visualization of data in a pie graph (or proportional representation) can lead them to develop intuitions of the role of large sample sizes in reducing variability that occurs within and across samples.*

## 3.5 Modeling Situations with the Weight Tool

When modeling a situation, it is not uncommon to observe students using an additive approach to create an equivalence relation of different sets of weights for a situation (e.g., Drier, 2000a, Stohl & Tarr, 2002b). For example, if given a bag of marbles with 4 red and 6 green, students would initially believe that weights of 1 and 3 could model this situation since 6-4 = 3-1. As students gain more experience or are able to readily apply multiplicative reasoning, they correctly use weights proportional to the context or to a different set of weights. In addition, when asked to create an equivalent experiment including setting a new Weight Tool distribution using weights different from a previous experiment, students often readily alter the order in which the possible outcomes are entered in the Weight Tool but correctly maintain the proportionality.

Of real interest to us is that regardless of the level of maturity in weight equivalence reasoning, students appear to expect that equivalent theoretical distributions should give similar empirical results. Students often realize after collecting empirical data that their “equivalent” weights designed with additive reasoning do not correctly model the context. For example, as discussed in Stohl and Tarr (2002b), Brandon and Manuel (age 11) used the Weight Tool to create a model (Pink: 4, Yellow: 2, Blue: 4) for a spinner with pink and blue sectors each 40% and the yellow sector of 20%. They subsequently ran several trials to collect data from their spinner experiment, used the pie graph as a primary representation to analyze data and test the “goodness” of their model, and used decimals and percents in the data table as secondary representations. They most often ran multiple sets of 100 trials and occasionally a larger number of trials. After Brandon and Manuel were convinced that their 4:2:4 model was accurate, a teacher challenged them to design a model of the spinner using a total weight of 50. Manuel typed in 20:10:20 in the Weight Tool. Brandon claimed, “that’s not right” and Manuel said, “I bet you a billion dollars it is.” The teacher–researcher asked Manuel to convince Brandon that 20:10:20 could be used to model the spinner. Manuel struggled to explain how the weight model was in proportion to the original weights of 4:2:4 or the spinner regions. Brandon decided to run simulations in *PE* to “see if it still comes close, as long as we have the same percentages.” He first ran 100 trials with the pie graph and data table open and after 60 trials said, “That looks pretty right.” When the 100 trials were complete [showing a 34:23:43 distribution] he said, “Okay, that’s right.” Brandon continued to run sets of 50 and 100 trials and compare the pie graph image and percents to the physical spinner. Brandon compared percentages of the theoretical distribution in the Weight Tool with the empirical data shown in the pie graph and data table. He used empirical data to support or confute the notion that weights of 20:10:20 appropriately modeled the spinner. It is important to note that Brandon needed to run several sets of 50 and 100 trials before he was convinced that the empirical data supported the use of weights 20:10:20 to model the spinner.

> *A key lesson learned is that students seek relatively stable and similar repeated empirical results, in graphical and numerical form, as a means of comparing the goodness of a model and for comparing the equivalence of models.*

## 3.6 What Does a Weight of Zero Do?

Since the Weight Tool currently uses whole numbers to model a distribution, students are often faced with situations when zero is used. Some students seem to be able to connect this with empirical results where that outcome does not, or could not occur. For example, as discussed in Drier (2000a), Jasmine had designed an experiment with four different icons (tails, circle, hexagon, volleyball) and equiprobable weights of 1. When asked if they each had the same chance, she used the Weight Tool to illustrate her thinking.

Jasmine: They each have one. But they wouldn’t have the same chance if someone did [she changed the one under the tails to a zero] that. Then there would never be any of those [tails]. Or how about this? [she changed the zero under tails to be a two] now it’s more likely to get the tails because there are two out of five. But there’s only one circle, one hexagon, one volleyball out of five.

However, many students do *not* initially interpret a weight of zero as meaning the associated outcome is absolutely impossible, just highly unlikely. These students willingly run relatively large samples in search of this outcome occurring and are surprised when it does not happen. For example, in Drier (2000c), Dean and Lydia (8 years old) were modeling a situation where they were choosing whether to play soccer or baseball. When asked to design the chances so they were *certain* to play soccer no matter how many times they ran the experiment, they gave the soccer ball a weight of 12 and the baseball a weight of 0. Looking over at their screen, Jon turned to his computer and used weights of 19 and 0. When asked if the different computers would give different results since they used different weights, Jon thought his weight of 19 made him "more certain" to get a soccer ball than the other computer. Dean promptly said "it doesn't matter since we both gave baseball 0." After 100 trials and all soccer balls, Lydia was surprised and then, after a pause, noted "it doesn't matter what number you use as long as you give it all to the soccer ball." Similarly, in Drier (2000a) Amanda thought that if she used one and zero as the weights for heads and tails in a coin toss, there would be more heads, but that a "few tails" could occur. She ran a simulation with these weights and after about 400 trials decided that tails would not occur because "it's like there are none in the bucket." It was actually this comment, and Amanda's struggle to interpret the representations in the weight tool, that inspired the creation of a dynamic link between the weight tool and the bag of marbles.

Some of the students' initial interpretations of a weight of "0" demonstrate that the number zero itself may be too abstract to interpret in a meaningful way. One possible interpretation is that students think that the mere listing of the outcome in the Weight Tool asserts its possibility. But how is this possibility coordinated with a weight of zero? What if students are envisioning the weights as the "usual" distribution of outcomes in a sample of size equal to the total weight, but being random means that there is some variation in these sets. It is in these variations where any outcome listed can occur. So a zero weight is interpreted as the outcome not occurring in the "usual" sample, but the mere listing of the outcome makes it a candidate for appearance in the random "errors". Note that this would mean a subtle but real separation of probability and variation.

It is also worth noting that when building the distribution in the Weight Tool via placing marbles in a bag, students have never been observed in believing, say, a yellow marble will possibly occur if they do not place a yellow marble in the bag, even though a yellow marble is listed with a zero probability in the weight tool after construction. So whatever students are thinking, it seems to change if they construct weights via a more physical representation.

> *A key lesson learned is that a weight of zero is not directly interpreted as an impossible outcome, although a physical representation (bag of marbles) seems to eliminate this difficulty. In addition, the expectation that an outcome with zero as a weight can occur with a small chance may be related to students' imagination, and expectation, of a hypothetical experiment with results similar, but not exactly, like a probability distribution.*

# 4. APPLYING LESSONS LEARNED TO THEORY AND INSTRUCTION

In considering the results of the various studies with *Probability Explorer*, as well the work of several others (e.g., Pratt, 2000; Konold & Kazak, 2008) on the connections between theoretical and empirical probability, we propose a model that describes a bidirectional relationship between a probability distribution as a model for a probability context and empirical data from the same probability context (Figure 3). In addition, we also offer several implications from our work for instruction on probability.

## 4.1 A Model for Bidirectional Reasoning

The explication of the model we propose below is a refinement of that previously described in Stohl and Tarr (2002b), Lee, Tarr, and Powell (2005), and Lee, Angotti, and Tarr (under review). The lessons learned, as described above, from the cross study analysis of students using *Probability Explorer* suggests to us that this model may be viable for designing instruction and helping students develop an understanding of the connections between data and theory, and theory and data, in a robust bi-directional way.

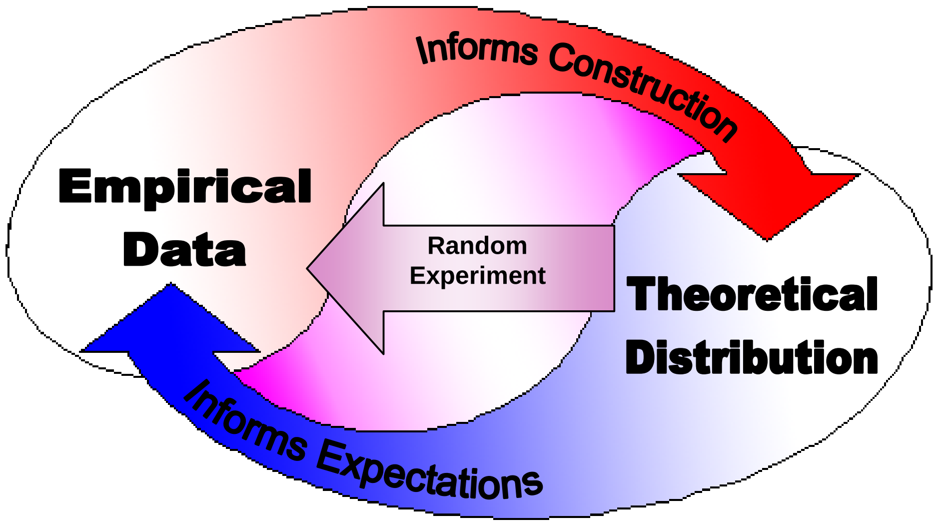


**Figure 3:** Model of bidirectional reasoning about probabilistic phenomena

Students may start from a theoretical model if they are able to well estimate a probability distribution based on known information in the probability context (e.g., rolling a six-sided die). In this case, their initial assumptions of a probability distribution inform an image of expected aspects of empirical data. After data collection, they may reflect upon the results against their current model of the underlying theoretical distribution (Konold & Kazak, 2008; Author 1 & colleagues, under review; Watson & Kelly, 2004). Noticing patterns in the data may make them call into question the prior assumptions in their model, or they may not believe the data varies enough from their mental image to contradict their initial assumptions. Their reasoning may then lead them to decide to collect and analyze more data to again test the reasonableness of the match between their hypothesized theoretical model and results from repeated trials. Note that it may be quite desirable to take an instructional approach to promote iterations of this possible cycle, as described in Konold and Kazak (2008) and used in Stohl and Tarr (2002b) and Lee, Angotti, and Tarr (under review).

Our model also promotes starting from empirical data when little is known about a theoretical model for the probability context. Students may conduct an experiment (or observe data) and wonder about the probability of an observed event where they have no prior experience with the phenomena or cannot use a classical approach to build a theoretical model based on a sample space (e.g., how likely is it that a tack will land on its side when dropped on the floor). Thus, they may start by conducting experiments, examining empirical data and using the relative frequencies from that data to inform the construction of a theoretical model, including a probability distribution. The first hypothesis about a theoretical model allows students to form a mental image of the expected results in future data. Students may also use their model to inform how (or whether) to collect more empirical data about the phenomena. Investigative cycles in this direction also seem worthy of promotion, and have been used in the instructional sequence described by Stohl and Tarr (2002b).

The robustness of students' reasoning from empirical data back to a probability distribution is influenced by their attention to sample size, understanding the independence of trials, and attending to variability of their data. Students need to consider that different trials and different sets of trials (samples) are independent of one another and variability among individual trials and samples is to be expected. They also need to coordinate conceptions of independence and variability with the role of sample size in the design of data collection and interpretation of results. Relative frequencies from larger samples are likely to be more representative of the theoretical distribution (regardless of their internal representation as physical or informational attributes) while smaller samples may offer more variability and be less representative.

## 4.2 Implications for Instruction

In order to promote students' bidirectional reasoning between theoretical distributions and empirical data, instruction should include opportunities to engage in tasks that allow them to reason about theoretical distributions that are known or are intuitively understood (i.e., equiprobable ones) as well as those that are unknown. In the first type of task, a goal for student is to explicitly compare empirical data distributions to that of the theoretical distribution and to notice conditions under which the two distributions appear more similar. In the second type of task, students also engage in collecting and examining data distributions, but the goal is to make an inference about an unknown theoretical distribution. While starting with the first type of task seems essential, the second type of task is needed to develop informal inference skills that can eventually support formal statistical inference. The first kind of task informs students as to the characteristics of samples that can support their reasoning in making inferences in the second task type. Coupled with the use of simulation technology, problem tasks similar to those used by Stohl and Tarr (2002) and promoted by Zieffler, Garfield, delMas, and Reading (2007) may offer students opportunities to grapple with numerous issues central to the study and understanding of informal inference. In so doing, students can learn the value of formulating data-based arguments and recognizing different characteristics of samples over a variety of sample sizes and theoretical distributions.

While it may be helpful at times for a teacher to suggest a sample size when students are conducting a simulation, we have learned that students can discover powerful ideas when given the opportunity to choose their own sample sizes. This freedom of choice needs to be followed, however, with a demand by the teacher for a rationale to support such choices and whole class discussions in which students can compare results with various sample sizes. Such reasoning and discussions can be useful in helping students understand the strong influences of sample size and characteristics of the theoretical distribution upon the variability among the empirical distributions and the variability between the empirical distributions and the theoretical distribution. An understanding of these relationships can further inform students' expectations in empirical distributions based on what they know about the theoretical distribution and sample size.

Probability tasks provide opportunities to build towards and upon students' ability to reason proportionally. The pie graph and data table (decimals, fractions, and percents) are useful representations of an empirical distribution that can facilitate proportional reasoning, both during experimentation, and in static form after data collection is complete. We suggest taking advantage of students' natural tendency to use the pie graph representation because of the way it "wiggles" (Carmella, age 9; Drier, 2000a) during data collection. The dynamic change of the graph as the sample grows is a way to focus students' attention to the correlation between the decrease in variation in proportions and the increase in sample size. In addition, we suggest capitalizing on the display of percents in the current Weight Tool and the data table as an explicit way to draw students' attention to the differences between the theoretical and empirical distributions.

# 5. REFLECTIONS ON DESIGN IMPROVEMENTS FOR BIDIRECTIONAL REASONING

As mentioned earlier, several other researchers have successfully designed software tools and documented how students are able to make connections between distribution of data from a random experiment and the theoretical distribution described in the model (e.g., Abrahamson & Wilensky, 2007; Konold, et al, 2007; Konold & Kazak, 2008; Pratt, 2000, 2005). Collectively, the results from these studies, along with the results from the various studies using *Probability Explorer* described in Section 3, can inform the next iterative cycle of software design. Although many improvements have been planned based on the results from the collective studies, the focus here will be on one proposed change. Specifically, we are considering what Weight Tool redesigns may better help students coordinate representations of the probability distribution with those of the empirical results that can promote bidirectional reasoning about probabilistic phenomena.

The first response may be to keep the Weight Tool essentially the same, but to add on the ability to view the theoretical distribution in the same bar or pie chart as is available for the empirical results. Being able to watch a static pie chart of the theoretical distribution beside the wiggly, slowly stabilizing pie chart of results from a running simulation would be quite useful to a student in solidifying connections between the two via some governing law of large numbers. However, many of the themes discussed above may be artifacts of the specific design of the Weight Tool as a representation of a theoretical probability distribution using whole number weights and a total weight.

Since the bag of marbles has shown promise in being a powerful metaphor, we would like to draw from the strengths of that, but we want students to move beyond (or avoid altogether) a total weight approach to their experimentation. Thus, we would need so many marbles that counting the total would be impossible, even as we maintain the appropriate ratios of subset cardinalities. Not only would this approach hopefully discourage the thought that a particular sample size is optimal, this approach may even suggest the desirability of a large sample size. On the other hand, some students may just take on the strategy of picking a sample size equal to a multiple of the number of outcomes.

The current Weight Tool was designed based on research of young students' tendency to relate probability in terms of part-to-part relationships rather than part-to-whole. Thus, an initial conjecture for the design was to have weights displayed and entered in part-to-part format but also displayed in part-to-whole format. This design can, and has been shown to, foster student's proportional reasoning. While that early design was built to help young students where they may be starting, we also need to consider a design for the Weight Tool that is robust enough to lead students where they are going—inferential statistics. An issue of particular importance to statistics is developing a successful transition from different representations (relative frequency tables, etc.) of discrete sample spaces to those (density functions, etc.) of continuous spaces (Lee, 1999). So, we have been pondering a critical question-- can this be done while maintaining intuitions behind the classical and frequentist approaches and developing and using students' understanding of proportions?

# 6. GRANULAR APPROACH TO PROBABILITY DISTRIBUTIONS

## 6.1 A New Paradigm for Modeling Probability Distributions

Introductory probability lessons typically attempt to build from at least one of two primary student intuitions: 1) the concept of equiprobable outcomes, 2) and the law of large numbers (Batanero, Henry, & Parzysz, 2005). Breaking sample spaces into equiprobable outcomes and viewing the probability of an event as a part-to-whole proportion is the primary paradigm known as the classical approach. This approach relies heavily on an intuitive understanding of uniform probabilities. The combinatorial distributions that can arise from this approach give a sound insight to the nature of population distributions found in the real world. Unfortunately, even the most basic combinatorics can be quite daunting to most students, not to mention the limiting processes involved with interpreting continuous distributions. In contrast, a frequentist approach, which often starts from the empirical side of the bidirectional model, appeals to the law of large numbers by describing the probability of an event as being a limiting proportion of a large number of quasi-identical trials. This approach gives little guidance to the interpretation of probability for a single or a small numbers of trials, nor few easy guarantees for relative frequencies in the long run. However, this approach binds well with the bidirectional model and the related investigative iterations.

In upgrading *Probability Explorer*, we propose a *granular density* paradigm to replace the integer Weight Tool. A unit amount of "sand" will contain grains, each grain being equally likely to be selected in a trial. Any quantity of sand would contain a "large" number of these grains. Thus, both primary concepts of equiprobable outcomes and law of large numbers are integrated in all scenarios.

Let's consider the proposed design for representing the theoretical probability distribution in *Probability Explorer*. As a student decides on the *n* possible outcomes for the sample space, the new Distribution Tool will default to *n* bins of equal width. The content of each bin represents the probability for that outcome. A unit volume of sand will be in a large container with the goal of redistributing it to the bins below to assign a probability to each outcome. The user has the option of auto-distributing the sand equally to all bins (to quickly model equiprobable contexts such a fair six sided die roll) or manually "pouring" an amount of sand into each bin (Figure 4a), either through typing in numerical values or manually operating the spigot.

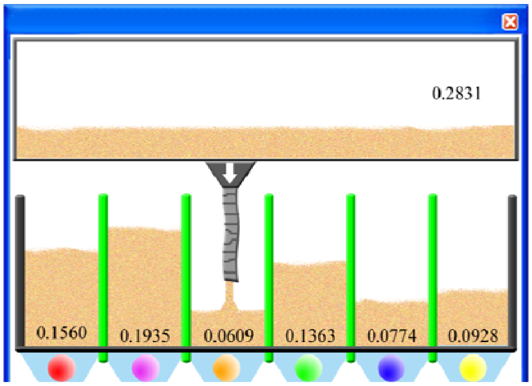


**Figure 4a:** Redistributing sand into bins

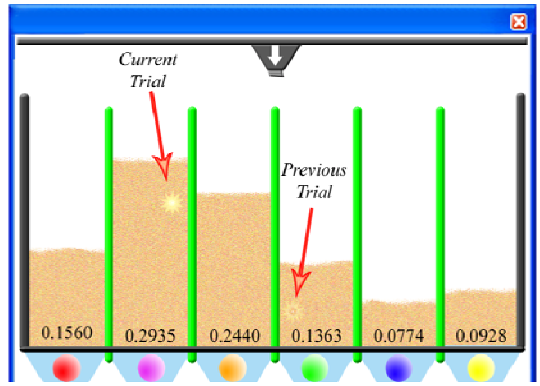


**Figure 4b:** Illuminated grains of sand

In the current design of the Weight Tool, there is not a clear dynamic connection between the Weight Tool and the process of generating a random trial. In fact, most of the time students close the Weight Tool when running trials, and do not often have a visual reminder of the theoretical distribution on the screen while they collect data and analyze visuals of the dynamically growing empirical distribution. In the new design, if the Distribution Tool remains visible when a user presses the Run button to execute a number of trials, for each trial, a grain of sand will be randomly picked and illuminated (Figure 4b). The trial outcome that occurs corresponds to the tab icon for the bin containing the illuminated grain of sand (in Figures 4-6 the tabs are various colored marbles). This feature should allow for a better visualization of how the empirical data is generated from and connected to the probability distribution.

To build off the powerful metaphor of “marbles in a bag” and to hopefully avoid misinterpretations of a weight of zero, the unit of sand can also be more grossly discretized into “bricks” that can be distributed to the bins by either dragging a brick, clicking on the outcome icon to auto-move a brick, or typing in a numerical value to indicate the quantity of bricks to place in an outcome bin. In order to stay true to our desires to give students access to the ability to model a probability distribution in a part-to-part manner, the bricks allow a user to conceive of the total sand as divisible into a small number of equal parts. The redistribution of the bricks of sand then becomes a partitioning task of sharing *r* bricks among *n* outcomes (Figure 5).

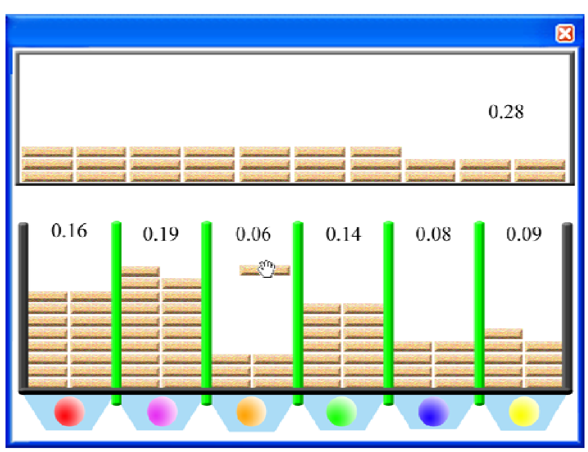


**Figure 5:** Redistributing sand as bricks

Of course, we recognize the potential for *n* bricks to also promote a total weight approach to experiments. However, the bricks can be "broken" back into sand grains to help transition from a discrete to more continuous model. Thus, whether the sand is contained in "bricks" or not, the entire probability distribution is made up of a very large, un-named number of tiny grains of sand. We believe this large number will reduce the total weight approach in students' data collection and may be suggestive of collecting large sample sizes.

Until a student places sand into a bin, the probability will be displayed as zero. This can hopefully help students conceive of a way of correctly interpreting a probability of zero. We are also considering whether or not we should require all sand from the unit to be distributed before a sample is taken. Allowing some sand to remain in the unit container could have interesting implications and uses for conceiving of conditional probability.

There are times when an instructional task is purposely designed such that students are starting from the empirical side of the bidirectional model. In such a case, there needs to be a way for a teacher to hide the underlying theoretical distribution that is governing the pseudorandom number generator from the students. This capability is in the current version of *Probability Explorer*, and will remain possible in the new version. However, in the new version, students will have the opportunity to use the Distribution Tool to distribute sand into bins to design a distribution to represent their current hypothesis for a model of a probability distribution, based on their empirical results. Then they can revisit this hypothetical distribution and make changes to their model based on further data collection from the hidden theoretical distribution, *or* they could take and compare samples from the hypothetical distribution. This feature can help make *Probability Explorer* a tool that could allow teachers and students to engage in tasks that are aligned with our bidirectional model.

While there has been use of *Probability Explorer* as a tool for advanced statistics students in secondary school (e.g., Rider & Lee, 2006), the current design does not necessarily support advanced conceptualization of a continuous distribution and area under the curve as representing probability. As students become more sophisticated in their use of the sand bins, they will be able to drag the bin dividers to adjust the width of a bin. This can be done prior to redistributing sand from the unit container or after sand has been poured. In the later case, students may notice that the amount of sand in each bin is invariant (compare probabilities in Figures 4b and 6). We conjecture that having the bins being adjustable in width can strengthen the notion of probability being stored in area with height being a byproduct of density. This step from the discrete to the continuous, with this move from storing distributions as a finite set of numbers to storing them in density functions, is quite weak in most curricula. We conjecture that this granular density paradigm can provide a nice step in this transition. Further design and research will help us know.

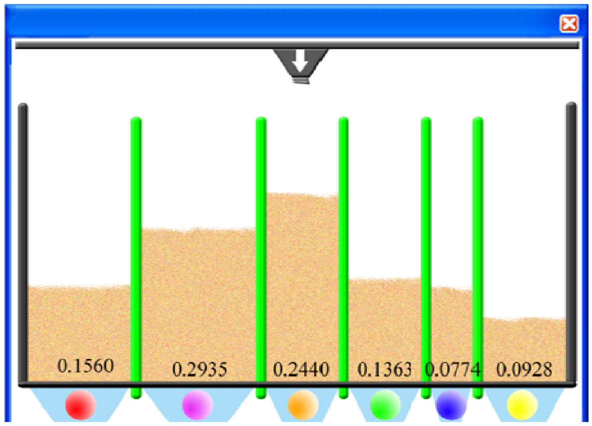

**Figure 6:** Adjusted bin widths

## 6.2 Future Research Directions

As we put our designs and ideas into action, it is necessary for us to conduct several different iterations of research. First, as we enact initial draft designs it will be necessary to conduct some usability tests with small groups of students to examine how they are interpreting the representations in the new Distribution Tool and they ways in which they naturally interact with the tool as they design and conduct simulations. This small scale research will inform revisions and the next iteration of designs. We recognize the potential of this first phase of research to drastically alter how we conceptualize the way we enact a granular density paradigm in the Distribution Tool interface.

Because the Distribution Tool is hypothesized to help students from a broad range of ages and grades (upper elementary through college level) develop an understanding of probability that is useful for inferential statistics, we will need to design several design-based research experiments. In these experiments we can hypothesize, enact, and examine different instructional sequences that can be used to create a more well-defined learning progression that could lead to students actively drawing upon robust bidirectional reasoning between theoretical distributions and empirical data.

# 7. CONCLUDING REMARKS

The use of simulation tools to teach probability and statistical concepts has increased over the past 20 years as teachers, designers, and researchers consider ways to harness the computing and graphical power of a variety of technological tools. What we have shared is a glimpse into the journey of one such tool and how we envision its future. The iterative nature of the theorizing, design, development, and research stages (Clements, 2007) that have occurred over the past decade with this tool have led us to a current state where we are ready to implement new improvements, including a paradigm shift in how distributions can be modeled in the software, and to further research the interactions between students' reasoning and their use of the tool. As suggested by Clements (2007), the processes used in educational software and curricular design need to be made more public and explicit for the benefit of other researchers and designers in the field. We hope this paper provides insight into our processes and goals and provokes thought and discussion on how our model of bidirectional reasoning and software designs may afford or constrain students' probabilistic reasoning.

## References


Abrahamson, D., & Wilensky, U. (2002). ProbLab: A suite of models for experiencing and relating micro and macro perspectives of probability. http://ccl.northwestern.edu/ProbLab/index.shtml Retrieved January 25, 2008.

Abrahamson, D., & Wilensky, U. (2007). Learning axes and bridging tools in a technology-based design for statistics. *International Journal of Computers for Mathematics Learning, 12*(1), 23-55.

Batanero, C., Henry, M., & Parzysz, B. (2005).The nature of chance and probability. In G. A. Jones (Ed.) *Exploring probability in school: Challenges for teaching and learning,* (pp. 15-37). Netherlands: Kluwer Academic Publishers.

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning 2*(1&2), 127-155.

Biehler, R. (1991). Computers in probability education. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: Probability in education,* (pp.169-211). Boston: Kluwer Academic Publishers.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student leanring of statistics. *Technology Innovations in Statistics Education 1*(1). Available at http://repositories.cdlib.org/uclastat/cts/tise/vol1/iss1/art2/.

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula". *Journal for Research in Mathematics Education 38*(1), 35-70.

College Board. (2007). AP Program Size and Increments. Available at apcentral.collegeboard.com/apc/public/repository/2007_Size_and_Increment.pdf

College Board. (2006). AP Statistics Course Description. Available at www.collegeboard.com/prod_downloads/ap/students/statistics/ap-stats-0607.pdf

Drier, H. S. (2000a). *Children's probabilistic reasoning with a computer microworld*. Unpublished doctoral dissertation, University of Virginia.

Drier, H. S. (2000b). Children's meaning-making activity with dynamic multiple representations in a probability microworld. In M. Fernandez (Ed.), *Proceedings of the 22nd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, (vol 2, pp.691-696). Tucson, AZ.

Drier, H. S. (2000c). The Probability Explorer: A research-based microworld to enhance children's intuitive understandings of chance and data. *Focus on Learning Problems in Mathematics 22*(3-4), 165-178.

Drier, H. S. (2001). Conceptualization and design of Probability Explorer: A research-based journey towards innovative educational software. *Tech Trends 45*(2), 22-24.

Franklin, C., Kader, G., Mewborn, D. S., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2005). *A curriculum framework for K-12 statistics education.* GAISE report. American Statistical Association. Retrieved September 3, 2006 from www.amstat.org/education/gaise

Jiang, Z. (1992). *Students' learning of introductory probability in a mathematical microworld.* Unpublished doctoral dissertation, University of Georgia, Athens, GA.

Jones, G. A. (2005). *Exploring probability in school: Challenges for teaching and learning*. Netherlands: Kluwer Academic Publishers.

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp.909-955). Charlotte, NC: Information Age Publishing.

Konold, C. (1992). *Prob Sim*. Software distributed at www.umass.edu/srri/serg/software/probsim.html.

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning 12*(3), 217-230.

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. Technology Innovation in Statistics Education 2(1). Online serial http://repositories.cdlib.org/uclastat/cts/tise/vol2/iss1/art1/.

Lee, H. S. (2005). Students' reasoning with small and large trials in probability simulations. In S. Wilson (Ed.), *Proceedings of the Twenty-Seventh Annual Meeting of the Psychology of Mathematics Education-North American Chapter*. Blacksburg, VA, October.

Lee, H. S., & Lee, J. T. (2009). Students' interpretations of probability distributions in a simulation environment. *Proceedings of the Twenty-Ninth Annual Meeting of the Psychology in Mathematics Education-North American Chapter*. Atlanta, GA, October.

Lee, H. S., Angotti, R., & Tarr, J. E. (under review). Making comparisons between observed data and expected outcomes: Students' informal hypothesis testing with probability simulation tools.

Lee, H. S., Tarr, J. E., & Powell, A. (2005). Working group for the complexity of learning to reason probabilistically. *Proceedings of the Twenty-Seventh Annual Meeting of the Psychology in Mathematics Education-North American Chapter*. Roanoke, VA, October.

Lee, J. T. (1999). Discuss with your colleagues: It's all in the area. *Mathematics Teacher 92*(8), 670-675.

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

Parzysz, B. (2003). From frequency to probability. Some questions posed by the new French senior high school curricula. Paper presented at the *54th International Statistical Institute*, Berlin, Germany. Retrieved March 1, 2005 from: http://www.stat.auckland.ac.nz/~iase/publications.php?show=3.

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. *The Journal of Mathematical Behavior, 22*(4), p. 405-435.

Pratt, D. (1998). *ChanceMaker*. Software available at http://fcis1.wie.warwick.ac.uk/~dave_pratt/page10.html.

Pratt, D. (2000). Making sense of the total of two dice. *Journal of Research in Mathematics Education, 31*, 602-625.

Pratt, D. (2005). How do teachers foster students' understanding of probability? In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 171-190). Kluwer Academic Publishers.

Rider, R. L., & Lee, H. S. (2006). Differences in students' use of computer simulation tools and reasoning about empirical data and theoretical distributions. *Proceedings of the 7th International Conference on Teaching Statistics*, Salvador, Brazil, July.

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

Shaughnessy, J. M. (2003). Research on students' understanding of probability. In J. Kilpatrick, W. G. Martin, & D. Schifter, (Eds.), *A research companion to Principles and Standards for School Mathematics.* Reston, VA: NCTM.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 957-1009). Charlotte, NC: Information Age Publishing

Stohl, H. (1999-2002). *Probability Explorer v. 2.01*. Software application available at http://www.probexplorer.com. Published by Author.

Stohl, H. & Tarr, J. E. (2002a). Using multi-representational computer tools to make sense of inference. In, D. Mewborn (Ed.), *Proceedings of the twenty-fourth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (vol 3, pp. 1317-1330) Athens, GA.

Stohl, H., & Tarr, J. E. (2002b). Developing notions of inference with probability simulation tools. *Journal of Mathematical Behavior 21*(3), 319-337.

Tarr, J. E., Lee, H. S., & Rider, R. (2006). When data and chance collide: Drawing inferences from empirical data. In G. Burrill (Ed.), *Thinking and reasoning with data and chance: 2006 yearbook of the NCTM* (pp. 139-149). Reston: VA: National Council of Teachers of Mathematics.

Tversky, A, & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 2,* 105-110.

Tversky, A., & Kahneman, D. (1982). Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 3-22). New York: Cambridge University Press.

Vahey, P. (1998). *Promoting Student Understanding of Elementary Probability Using a Technology-Mediated Inquiry Environment.* Unpublished doctoral dissertation, University of California-Berkeley, Berkeley, CA.

Watson, J. M., & Kelly, B. A. (2004). Statistical variation in a chance setting: A two year study. *Educational Studies of Mathematics 57*(1), 121-144.

Weber, K., Maher, C., Powell, A., & Lee, H. S. (2008). Learning opportunities from group discussions: Warrants become the objects of debate. *Educational Studies in Mathematics 68*, 247-261.

Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2007). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal 7*(2), 40-58.