Int J Comput Math Learning (2007) 12:217–230
DOI 10.1007/s10758-007-9123-1

Understanding Distributions by Modeling Them

Cliff Konold · Anthony Harradine · Sibel Kazak

Published online: 27 November 2007

Abstract In current curriculum materials for middle school students in the US, data and chance are considered as separate topics. They are then ideally brought together in the minds of high school or university students when they learn about statistical inference. In recent studies we have been attempting to build connections between data and chance in the middle school by using a modeling approach made possible by new software capabilities that will be part of *TinkerPlots* 2.0 (*TinkerPlots* is published by *Key Curriculum Press* and has been developed with grants from the National Science Foundation (ESI-9818946, REC-0337675, ESI-0454754). Opinions expressed here are our own and not necessarily those of the Foundation.). Using a new Sampler object, students build "factories" to model not only prototypical chance events, but also distributions of measurement errors and of heights of people. We provide the rationale for having students model a wide range of phenomena using a single software tool and describe how we are using this capability to help young students develop a robust, statistical perspective.

Keywords Probability · Modeling · Computer simulations · Statistical reasoning · Distributions · Data analysis · Co-variation

What do the sum of two dice and the length of cats' tails have in common? In this paper, we describe how we are involving middle school students in investigations designed to draw connections between prototypical chance events, such as the sum of two dice, and distributions of data, such as the lengths of cats' tails or the heights of people. In these situations, we would like students to come to understand that the reason values in the middle of the distributions are more frequent is because there are more ways to get them.

This work is being done as part of the Model Chance project, a 3-year project funded by the National Science Foundation to develop software and instructional activities for

C. Konold (✉) · S. Kazak
University of Massachusetts Amherst, Amherst, MA 01003, USA
e-mail: konold@srri.umass.edu

A. Harradine
Prince Alfred College, Adelaide, Australia

teaching data and chance in the middle school. The project aims to address limitations in existing US curriculum materials. Existing materials convey little sense of the importance and range of applications of probability, and prior to high school, they treat probability and data analysis as separate strands.

The project is developing an alternative approach that relies heavily on students using new simulation and modeling capabilities that we are developing as part of a future version of the data analysis tool *TinkerPlots* (Konold and Miller 2004). Teaching probability using a modeling approach has a number of advantages. Through modeling, students encounter and use probability in virtually the same way as practitioners do, with the purpose of better understanding some real phenomenon. Additionally, this approach frequently requires that students articulate their informal theories about probability and then put them to the test. This has been shown to be a powerful technique in fostering conceptual change in the domain of probability (Drier 2000; Konold 1994; Pratt 2000) where people have strong prior conceptions, many of which are at odds with normative theory (Tversky and Kahneman 1974; Konold 1991; Lecoutre 1992). Finally, unlike the majority of probability problems students currently encounter in school settings, which are framed in terms of single trials, simulations generally focus on what happens in large samples, a perspective that several researchers have now suggested is critical for forming or eliciting normative expectations about probabilities (Gigerenzer 1994; Konold 1989; Saldanha and Thompson 2002; Sedlmeier and Gigerenzer 1997; Shaughnessy and Ciancetta 2002).

In the Model Chance project, we are using this simulation capability to go beyond modeling traditional probability contexts. In this article, we describe how students use the same simulation capabilities that they use to explore probability to build "factories" that manufacture datasets. Our hope is that the process of building and refining these data-producing factories will prompt students to take on increasingly sophisticated data-modeling objectives. In our initial attempts at involving students, which we report here, we have been interested in the degree to which students spontaneously engage with and respond to the modeling challenges associated with building factories. Our long-term object is to design tools and activities that help students develop a statistical view of data as aggregate collectives which include signatures of the processes that produce them, including stable (signal) and variable (noise) components (Konold and Pollatsek 2002).

Elsewhere we have described features of *TinkerPlots* designed to support the development of a statistical perspective (Konold 2007). A basic premise is that a statistical learning environment should allow students initially to operate on data in ways that make sense to them before receiving any formal training in statistics. Thus students use *TinkerPlots'* divider tool to locate central clumps in data distributions, an action that fits with students' tendencies to view numeric data as comprising three groups: low, medium, and high values (Konold et al. 2002). However, these basic tools contain in them suggestions or implications for more expert-like use. Thus, with support, students begin using the dividers tool used in conjunction with other marking devices to locate averages of data and indicate and quantify the spread around these averages (Konold et al. 2007). In this way, we see tools as more than ways to express thoughts we already have or to accomplish goals within our current range of expertise; tools also set the stage for and support new types of thinking and action (Konold and Lehrer in press).

Among the most significant contributions of Jim Kaput to mathematics education were his attempts to reconceive mathematics through the lens of technology, and to use the new representational means of those technologies to widen the avenues into mathematical understanding, making it accessible to everyone. He saw the computer playing a pivotal role in mathematics education, comparable to that played by Gutenberg's printing press, which produced then empowered a literate society (Shaffer and Kaput 1999). Thus in newly-emerging technologies, Jim saw the promise of "democratizing" mathematics, and he set out to do that, focusing in particular on algebra and calculus. Even more radically, Jim argued that we were witness to the dawn of a new form of cognition made possible by the fact that these technologies were not only expanding our brain's memory capacity; paper and pencil began serving that function long ago. He claimed that we are now using personal computers to extend our mind's processing capabilities, and that with that new facility, the nature of our thinking would adapt and change (Shaffer and Kaput 1999; Kaput et al. 2002).

Viewed from that lofty perch, our own recent explorations, which we describe here, are barely discernable. Perhaps for this reason, we tend to use Jim's work as a backdrop, a surface on which we can project our modest steps to see them as but part of a much grandeur and evolving scheme. This accords with our sense that Jim could perceive more meaning and significance in articles and technology tools we had produced than we, their authors, could.

1 Introducing Students to Data Factories

In this article, we focus primarily on recent work with 12 students in grades 6–8 (ages 12–14) in Holyoke, Massachusetts. These students participated in a yearlong after-school program focused on exploring probability. The public middle school these students attended serves about 400 students who are predominately from minority groups (76%), with 85% of the students qualifying for free or reduced-price lunch. Schools in the Holyoke School District have some of the lowest scores in Massachusetts on the high-stakes MCAS test. More recently we conducted a similar activity with 15 boys (age 12–13) in a private school in Adelaide, South Australia, but we mention those results only in passing.

We introduced students to the data factory idea using the Sampler shown in Fig. 1. The Sampler includes two mixers and two spinners, which operate independently. The Sampler is set to create 100 "cats" (repetitions = 100) with four attributes. These attributes correspond to the names of the four sampling devices in Fig. 1: *Name*, *Gender*, *EyeColor*, and *Length*. In a previous activity, students had analyzed a data set of 100 real cats, which included these same attributes along with several others. Pressing the Run button on the upper left causes the mixers and spinners in the Sampler to operate in turn, randomly drawing first a name, followed by a gender, eye color, and length. The case (cat) being created is depicted by the circle icon and the four spaces to its right, which start out blank. The case icon moves from left to right along the bottom of the window collecting an attribute value from each device. In the example in Fig. 1, the "cat" being created has been named "Hi," given a gender of "f," and "blue" eyes. It is about to receive its final value from the *Length* mixer.

After a cat is created in the Sampler, it is added to the growing collection. On the left of Fig. 2 is a table showing the values for the 22 cats created so far. The plot window on the right displays these 22 cats according to their *Length* (x-axis) and *Gender* (two different colors, which here appear as different values of grey).

After seeing a brief demonstration of the factory and how we constructed it, students were eager to build their own. The only constraints we imposed were that their objects could not be cats and should have from 3 to 6 attributes.

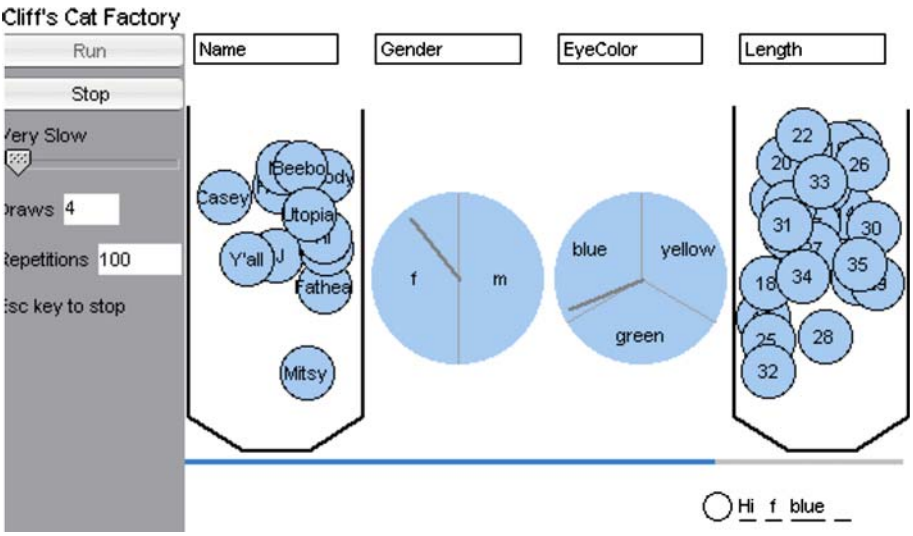

Fig. 1 A factory for making cats. It includes four sampling devices: two mixers and two spinners. Each device will determine the value of an attribute of a cat. The thicker line in the two spinners is the spinner arrow whose position after sampling determines the outcome of a repetition. The circle and four lines on the bottom right is a case under construction. Its values for Name (Hi), Gender (f), and EyeColor (blue) have been sampled, and it is about to receive a value for Length. As it is being constructed, the circle and four lines at the bottom move along from left to right

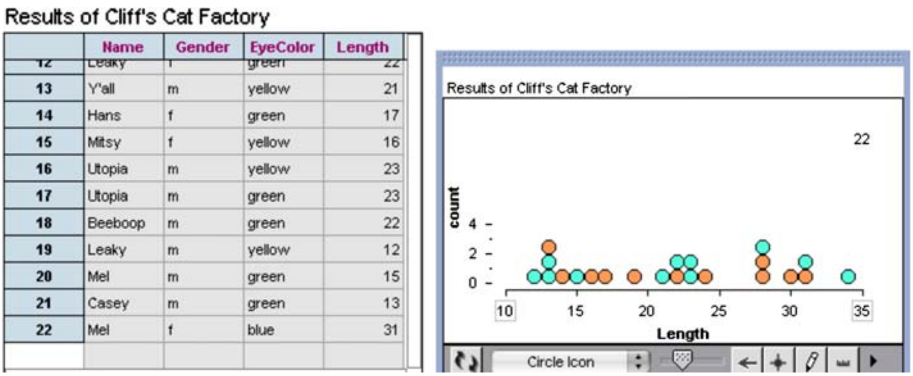

Fig. 2 After a case is created in the sampler, it appears in the bottom row of a case table (left). If a plot is open, the new case also appears in it. In this instance, the plot has been organized to display Length on the horizontal axis and Gender as two different icon colors (they appear here in grey scale)

2 The Challenge of Conceiving of Objects as Constellations of Attribute-value Pairs

The task of creating a data factory is more challenging than it might first appear, because it requires students to conceive of real-world objects as comprising a set of attributes according to which we can measure and classify them. The world is not populated by data—real objects are not bundled sets of instantiations of attribute values. We create data by coding and measuring selected features of real-world objects. Thus, conceiving of real-world objects as sources of data in this sense is the first step in modeling them statistically (Hancock et al. 1992; Konold and Higgins 2003; Lehrer and Schauble 2007).

Sitting down to develop their factory, one pair of eighth-grade girls asked, “What kinds of objects can we make?” Can we make candy?” to which we replied, “sure.” One of the pair happily wrote down, “We want to build a candy factory.” They were completing a worksheet that next asked them to list attributes of the objects from their factory. “Attributes. What do you mean, ‘attributes’?”

These students had previously been working with *TinkerPlots* to analyze pre-formed datasets. As part of those activities, students had been selecting attributes and using them to make graphs of the data to explore various questions. In these prior activities, students had learned to speak about and operate on data using attributes. But in these instances, attributes were always accompanied by cases with values on those attributes. Thus students had not yet needed to think about attributes, such as *Gender* or *Height*, in the absence of cases with values (cases with values of males and females for *Gender*, or 56 inches for *Height*). *Height*, in this sense, is an invented construct together with a set of measurement techniques and protocols that we use to assign unique values to cases. And the application of the attribute and accompanying protocols is more than a scheme for naming things; it is a system for placing cases in relation to one another with respect to that attribute (Bowker and Star 1999).

All of the students we have worked with to date have been able, with some thought, to come up with a reasonable set of objects and attributes to create. Perhaps influenced by our example, about 40% of the factories have produced people or animals. And in these instances the attributes have generally included gender and name, just as our example did. Other common contexts have included food and sports equipment (skateboards, surfboards), with attributes including brand, type, size, etc.

Students’ work was supported by a worksheet with written instructions that broke the task of factory creation into steps. After deciding on the objects, the worksheet asked them to generate a list of attributes for those objects. It seemed sensible to us that one would have to have in mind attributes before generating possible values. In retrospect, however, this was a simplistic view (and thanks to an anonymous reviewer for pointing this out). It seems equally plausible that in such a task students might first focus on some feature of an object, for example by imagining a cat with very long fur, and use that image to generate a contrasting case (a cat with short fur), and as part of that second act activate the relevant attribute “fur length.” We find some support for this account in the completed worksheet of a student who was designing a “Food products” factory (see Fig. 3). For the first three items he generated, he listed values (e.g., “creamy, crispy”) not attributes (e.g., “Texture”). It was only when he turned the page and was asked to fill in attribute names on a “data card” (right of Fig. 3), that he explicitly named many of the attributes. In those

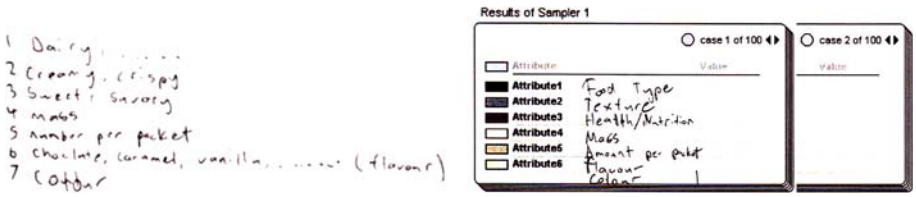

Fig. 3 Attribute names and value for a student’s “Food products” factory. The student first generated the numbered list at the left. Afterwards, he filled in attribute names on the card to the right. Thus he explicitly named several of the attributes only after he had generate a list of values

instances when what he first listed were values rather than attribute names, we speculate that the value names were more familiar nouns (e.g., "Dairy") than were the attribute names (e.g., "Food Type"). The only category attribute for which he first produced the attribute name was *Colour*, which is probably as commonly encountered a word as the first attribute value he listed: "Red." For both of his numeric attributes ("mass," "number per packet") he produced attribute names first, and never listed possible values. This may be related to the fact that for numeric attributes, values (such as 13) are not highly predictive of the attribute, whereas with category attributes, they typically are fairly predictive.

Rather than asking students to initially list attributes and then values, as we did, it may be more productive and revealing to first ask students to specify how their created objects will differ one from another. This would allow students to focus either on distinguishable attributes or on attribute values (or possibly both). We expect, however, that it will still pose a challenge to students to then construct these differences explicitly as attributes and their values.

A possible confusion in our task specification resulted from the "data factory" metaphor. Ryan and Nelson, who we videotaped as they planned and then built their factory, decided to make skateboards. Under the worksheet section asking what kind of objects they would make they wrote "Skateboards, trucks [axels], and wheels are the objects." When asked to list attributes of these objects, they were confused. But after inspecting the cat example, which was still visible on the screen, they concluded that what they had listed as the objects were in fact the attributes. When they then tried to specify values for the attributes, Ryan responded, " Value. What do you mean, value? How much they'd be [the cost]?

We speculate that what they had listed as attributes were not viewed by them as attributes on which skateboards could vary; rather they were the basic components from which you would build a skateboard—a board, trucks, and wheels. Our factory metaphor may have encouraged this interpretation, leading them to think about the parts they would need to assemble a skateboard. After more thought and inspection of the cat-factory example, Nelson was able to modify his perception, suggesting that for a value of "Skateboard" they write "Hawk" (a skateboard brand name). Later they added "Element," another brand name. In the row labeled "Trucks," Logan suggested "Steel" and later "Aluminum." Thus what we think they had been using previously as the name of a component common to all skateboards (Wheels), they had transformed into the name of an attribute on which skateboards varied ([Type of] Wheels).

For the most part, we believe the difficulties students experienced in this task were not due to superficial features of the task specification. Rather, the task problematized what, until this activity, had been a transparent structure. In analyzing data already entered into *TinkerPlots*, such as the dataset based on real cats, the screen objects behaved as "talking cats" might. Students could click on an attribute name, "Gender" and in so doing "pose a question" of the icons on the screen: "What's your Gender?" The icons instantly "responded" by assuming their corresponding color (gender value). There was no reason for students to consider how data might be structured such that they could "respond" in this way. To construct their factories, however, students had to conceive of the "questions" and the "answers" together as well as figure out how to express these two aspects in the data-structure syntax of "attributes and values."

There are other ways, of course, to problematize data structures. An interesting task that Hancock (1995) gave his students using a similar data tool, *Tabletop*, involved looking at a graph showing cases with names (Mary, John, Sally) separated into male and female bins. The students' task was to enter data into *Tabletop's* row and column database such that they could then generate the exact display. Most of the 13 students (aged 10–12) could not do this correctly in their first attempt, even though by that time they had gained considerable experience over 2 years both analyzing existing datasets and entering data themselves into *Tabletop*. Among other things, the task revealed that many students did not understand that the computer knew nothing about "names" and "genders" and thus did not know that "Mary," for example, was a female name.

The more common way to problematize data structure is to involve students in the creation of surveys to study some question of interest. Survey construction often involves students in resolving modeling issues that are bypassed in the data-factory task, such as the problem of translating real questions into statistical ones (Lehrer and Romberg 1996; Lehrer and Schauble 2007; Russell 2006). One of the drawbacks of survey construction is the time required. Too often, critical flaws are discovered after days of work, with no time left to resolve them (Hancock et al. 1992). In contrast, the data-factory approach allows for rapid cycles of testing and revision. Once on the computer, students could run their factory at any point, and we observed many of them modifying their factories after determining that it was not behaving as they expected. It was this feature of the activity, perhaps more than any other, that allowed the students we have worked with to create datasets within 30 min that seemed to roughly correspond to what they had set out to produce.

3 Populating Attributes with Reasonable Values

In introducing the data-factory activity, one question we have been interested in is whether students would enter attribute values that one might expect to observe in the real-world. For example, would students creating a people-making factory specify plausible values for height? To date, nearly all of the student-created attributes have included plausible ranges of values. For example, the girls making the candy factory included an attribute *Weight*. To decide values to include, they quizzed themselves and students around them about how many ounces certain types of candy might weigh, eventually concluding that weights of 1 to 3 ounces were reasonable. A student who made a people-making factory included the attribute *Age*, which ranged from 1 to 20, and *Height*, which ranged from 120 to 180 cm. The range of heights are certainly plausible for older people, but not for the younger ones.

Although the ranges of values students have entered have in general been reasonable, the distribution of values has not been: Students almost always included in their spinner or mixer device only one element of each value type. In the case of the people factory mentioned above, for example, the *Age* mixer contained one element for each of the ages 1–20; similarly, the *Height* mixer contained one value for each of the possible heights. Sampling from such factories, of course, tends to produce uniform distributions rather than non-uniform ones that we typically observe for the sorts of attributes students were creating.

There are several plausible explanations for this observation. Firstly, in our classroom experiments, we have only given students about an hour to create these factories. Given more time, many students might continue to fine-tune their models in a number of ways, and perhaps by attending to the frequencies of various values. Secondly, when students add additional elements into their sampling devices, by default the elements are equally likely. Creating a factory in which the device elements are not equally likely thus takes additional work. Also, we had not at this point involved the students in modeling problems in which the simple outcomes were not equally likely. These possibilities suggest that we might see different behavior if we simply had given students more time or had provided them experience in creating devices in which the elements were not equally likely.

But we believe that there are deeper reasons for the element structures of their final devices. Most of the students viewed the output from their factories not in graphs, but in data tables (as on the left side of Fig. 2). Because distributional characteristics of data are difficult to observe in tables, students may not have noticed that the output from their factories did not appear like other real-world data they had encountered. Thus, having students look at graphs of their data and consider whether their data were believable might prompt them to modify their models. Later in the article, we cite some evidence for this possibility. It may also be that students cannot yet anticipate how samples of data from a randomizing device will be distributed. Indeed, considerable research has shown that most students are not oriented to perceiving data as distributions as opposed to individual values or value-types (Ben-Zvi and Arcavi 2001; Khalil 2005; Konold et al. 2004). One of the questions we are exploring is whether modeling activities of this type might serve to develop distributional perceptions and reasoning.

Six months after the students in Holyoke had built these factories, we repeated the activity with two of the students in the after-school program. In the interim, they gained additional experience in using the software to make models that produced non-uniform distributions. On this second opportunity to build data factories, Byron made the computer factory shown in Fig. 4. It included two attributes, *Color* and *Size*, that were not symmetric. He explained that in his company, he'd make more black computers because "most computers are black." On this occasion, he did make a graph of his data (right side of Fig. 4). At the time he made the graph, the *Color* spinner was non-symmetric, as shown below, but the areas of the *Size* spinner elements were all equal. After inspecting the graph showing the distribution of sizes, Byron revised his *Size* spinner, making the sections unequal as shown in Fig. 4 because "people don't want to buy small [1–10] computers… and then also larger ones [41–50] are going to be really expensive, so I made that kind of

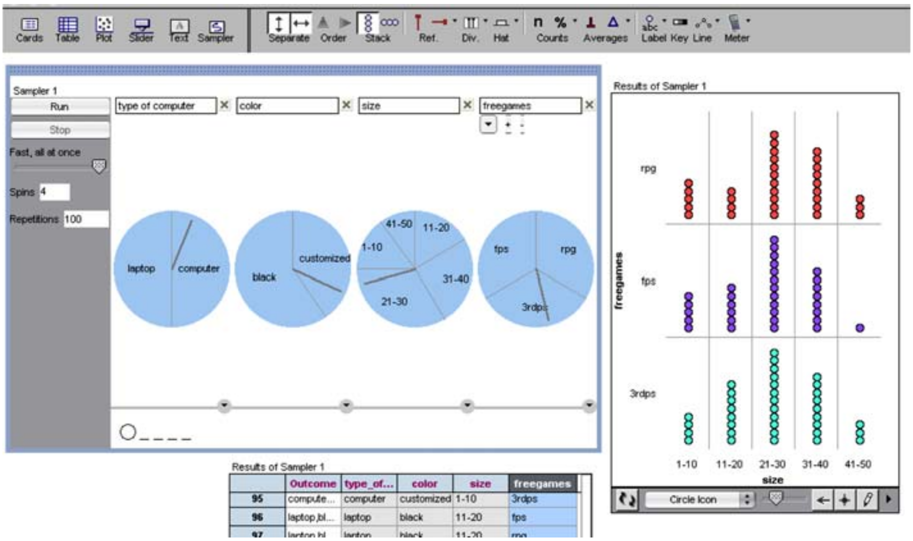

Fig. 4 Byron's computer factory and the data of 100 cases created from it. In this instance, Byron adjusted the area of the elements in the Color and Size spinners in accord with his expectations about consumer preferences. The spinner labeled "Type of Computer" includes "laptops" and "computer"[desktop]. The spinner "Freegames" includes three different types of computer games: "rpg" (role-playing game), "fps" (first person shooter), and "3rdps" (3rd person shooter)

small." Our speculation is that seeing the distribution of output from his factory cued his thinking about what the distribution of the size attribute should look like. In the next section, we will describe another student, Erin, who made similar adjustments only after looking at graphs of the output from her model.

We should stress that, though we were interested in whether students would strive to make realistic data, this was not an objective we gave them; they seemed to adopt this stance spontaneously. On the other hand, the stated task of designing a factory that makes real-world objects implicates modeling—what else could we mean by "real-world objects" than that they bear some resemblance to observable objects. However, in some cases students purposefully implemented designs to create unrealistic cases. The most common form of this playful turn was to use the names of classmates in a factory producing e.g., frogs or, in the case of the all-boys class, to create a device with names of classmates and another with males and females, and then watch to see who underwent a sex change. In our view, building factories that produce data that contrast in predictable ways with real-world observations make no fewer demands on the designer than building in correspondences.

4 Modeling Co-variation

Data factories, like those shown in Figs. 1 and 4, may create a set of attribute values that are distributed realistically, but the data as a whole would still not resemble a real dataset. This is because unlike real data, where there are dependencies between attributes, the values from the "linear" factories in Figs. 1 and 4 are drawn independently. In the cat factory, for example, males often end up with female names. When the Massachusetts students built their first data factories, the software did not yet allow other than linear arrangements of sampling devices. We were interested in whether students would notice the limitation this imposed.

After these students had built their factories, we had a brief class discussion. Students from two different groups described relations between attributes that they would have liked control over. Paul and Ken had made skateboards with attributes including *Length*, *Weight*, and *Brand*. They were disappointed that they could not represent differences between brands.

> Paul: Like if you're making skateboards, if you could match the brand with, like, what kind it is. Like, if it's a light skateboard, you could match it with the brand that makes those.

Similarly, Matt and Erin had made rollercoasters with attributes including *Name* (e.g., "The Monster," and "Kitty World") and *Size* ("big," "small"). They had wanted to

> Matt: … match, like, the name, 'cause we wanted to make the name like it would seem like a big rollercoaster. But it would come out as, like, small on the size.

This may be an example of where having a "bad" model may serve to highlight an important aspect of a situation that otherwise might go unnoticed. Our guess is that if you showed students rollercoaster data in which there was a dependency between name and drop height, for example, they would pay little or no attention to the dependencies between the two. But data in which those dependencies are absent are likely to attract attention, perhaps even more so when students view themselves as agents exercising some control over the data.

A few months after this first activity, we implemented a branching capability in the software that directs cases with different values on one attribute to different sampling devices (see Fig. 5). To introduce the students to it, we revisited the cat factory and pointed out the problem with how the linear factory was often misassigning males and female names. To remedy that, we built a new factory that first assigned values of *Gender*, and then branched males to one *Name* mixer with male names and females to another mixer with female names. Following this short introduction, we gave students three situations to model. Their instructions were to "Build a *TinkerPlots* factory that makes people that are:

(1) girls or boys, then gives them a hair length.
(2) girls or boys, then either pierces their ears or not.
(3) either wearing seat belts or not, then are either injured in an accident or not injured in an accident."

In the end, four of the five students attending that day were able to build reasonable models of these situations. Given that conditional probabilities and two-way tables are notoriously difficult for even college students to reason about (Batanero et al. 1996), the success of this "reversed" approach is promising. That is, although students traditionally have difficulty reasoning backwards from data to valid claims about conditional probability, these middle school students seemed quite able to reason forward from conditional situations they knew well to models and resultant data that represented those relations.

Figure 5 shows Nelson's model of the hair-length problem, which he described for us:

> Nelson: It will choose if it's male or female. And if it's male, it will go up here and get 1 to 5 inches of hair, and if it's female it will go down and choose 1 to 12 inches of hair.

He picked these values because "girls have longer hair and most … boys have real short hair." Even though he and three other students had little difficulty modeling a co-variation

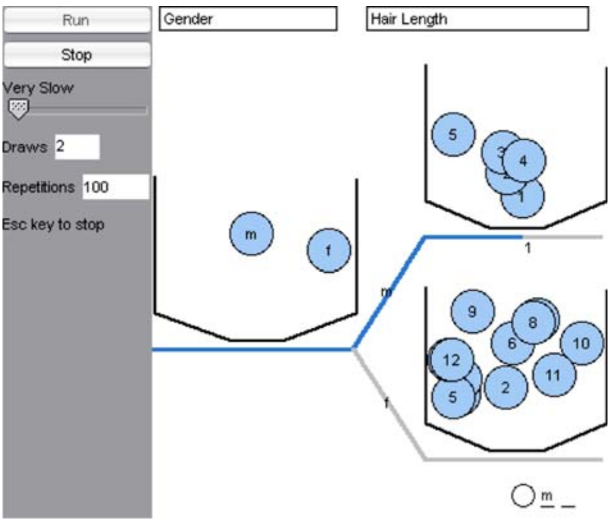

Fig. 5 Nelson's branched hair-length factory. We see a male case being produced, which has just received a hair length of 1 from the upper right mixer. Females go to the lower branch and receive a hair length of 1–12 inches

between gender and hair length, most of them continued to build devices where each value of hair length in a particular device was equally likely to be sampled. Perhaps because they could achieve perceptible difference between groups by using different ranges, there was no immediate need to attend to the distribution shapes. This approach fits, to an extent, with the tendency of students to compare two distributions (such as height of boys and girls) by noting which group has the highest (or lowest) values.

The exception was Erin, who rather than using numeric values for hair length used the categories "long," "short," and "medium" (see Fig. 6). Initially, she made identical hair-length spinners for males and females, with each length category equally likely. After producing 200 cases, she made a graph in which she organized the cases into six stacks of gender/hair-length, ordering them by gender, and within gender, by hair length. Looking at the graph, she observed:

> Erin: I see that it's more likely for a female to have—Wait now. I don't think that's really true… Because if you go walk around the streets of Holyoke, well actually anywhere, you'll see more males having short hair … The numbers won't match up.
> Instructor: So you want to do this more realistic?
> Erin: Yeah. How?

She had missed our demonstration in the previous session of how to freely adjust the boundaries of spinner slices. After we demonstrated this capability to her using a different context, she explained how she could make her hair-length factory more realistic.

> Erin: I could increase [the area of] the ones that seem more realistic [common], and decrease the ones that seem not that much realistic… Like, you don't see that many males with long hair…. Maybe if you go to California, but we're not in California.

That she critiqued the model only after she graphed output from it may suggest that at this point in the teaching experiment, distributional features of data were still tied for her to graphs. After seeing the data, she modified the spinner that determined male hair lengths, but left the female hair-length spinner as it was. We do not know whether she believed that

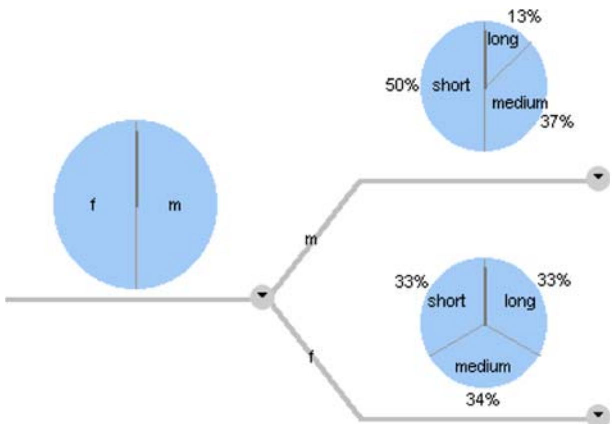

Fig. 6 Erin's hair-length factory after she adjusted the top right spinner to produce data more in accord with her expectations

the three lengths for females "on the streets of Holyoke" occurred equally often, or whether it was good enough in her mind to have captured the frequencies of male hair lengths while at the same time building into her model a difference between male and female hair lengths.

When she moved to the next problem (females and males with or without pierced ears), her initial model again branched from a gender spinner to two identical spinners which had equal probabilities for pierced and un-pierced ears. Thus her model again was set up to produce uniform distributions, despite the fact that she expected females to be more likely to have pierced ears. She again modified these in accord with her expectations after looking at output from her first model. This leads us to believe that she was, at this point, still unable to visualize what a distribution would look like by considering only the properties of the spinners, but could alter her model in sensible ways after seeing its output.

We should point out that the problems we gave students were worded to support their building appropriate models. The order in which the two attributes were introduced were consistent with the causal relation between the two. Also, the wording strongly suggested a two-stage process that would first produce an outcome on one attribute, and then an outcome on the second, conditioned attribute. Had our instructions instead been to build a factory that e.g., produced males and females with pierced and unpierced ears, there would have likely have been additional challenges for students. Some, for example, may have proceeded by creating combined elements (m-p, m-np, f-p, f-np) in a single device, which would then have required building in dependencies by manipulating the frequency of each pairing.

5 Conclusion

The results of our preliminary work with students building and running data factories have been encouraging. Students are quickly engaged with the task, approaching it initially as a playful act of creation. But without prompting, the activity quickly evolves into an act of modeling, as students want their created objects to look more like their real-world equivalents.

Viewed from the perspective of recent critiques of statistics instruction (Cobb 1993), the data-factory task may seem like a throwback from earlier textbook treatments in which data were typically treated stripped of context or dressed up in artificial or contrived ones. Lehrer and Schauble (2007), in fact, argue that things have not changed much, that "Although it may seem obvious that data only stand as data in relation to an issue of question, most data and statistics curricula place little or no emphasis on question posing" (p. 152). The data-factory task, in this respect, lacks many critical aspects of an authentic data exploration. Done in isolation it would be of no or limited value, educationally. The potential value we see in the task comes only as it serves our larger instructional objectives.

As we intimated in the introduction, those instructional objectives and the means for achieving them have been influenced by Jim Kaput's view of tools as being more than vehicles through which we can express existing ideas or accomplish currently-held objectives. Ideas and objectives are also prodded on by and reshaped through the tools we develop and use. Computer tools, in this respect, are not unique. But what today's mathematical and statistical computer tools do provide for students, and which is new, is the ability to obtain rapid feedback regarding the effects of tool use. In the case of Sampler, students can observe, seconds after making changes to their factory, the results of those changes in a graph. Actually, it is not quite this straightforward, as students must learn new ways to perceive data displayed in graphs before they can attend to the relevant aspects in the data (Konold et al. 2004). Indeed, one of our primary objectives in developing Sampler and housing it within a data tool is to support the development of these new, aggregate ways of seeing data. And as students develop these new perspectives on data, their use of the tools change (Konold and Lehrer in press.)

This and other modeling-based activities that we are classroom testing are ultimately aimed at helping students develop the foundations of statistical inference. These include activities where students build models of measurement error and natural variability that produce distributions similar to those obtained from real-world measurements (Lehrer et al. 2006; see also Wilensky 1997) as well as analyze output of models whose content they cannot see but must guess based on the data generated from them (Konold et al. 2007). Our conjecture is that the ability to create virtual data puts students in a more natural orientation vis-à-vis statistical thinking, allowing them first to attend to causal mechanisms (or parameters), and then afterwards to see how those mechanisms influence the resultant data. We expect that with sufficient experience operating on data from this perspective, students will come to view real data in the same way, i.e., as having been produced by a collection of independently acting mechanisms (or a parent population) which they can make conjectures about based on the data they have.

References

Batanero, C., Estepa, A., Godino, J. D., & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education, 27*, 151–169.

Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics, 45*, 35–65.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT press.

Cobb, G. (1993). Reconsidering statistics education: A National Science Foundation conference. *Journal of Statistics Education, 1*(1), Article 02. Retrieved from http://www.amstat.org/publications/jse/v1n1/cobb.html. Accessed 6 Nov 2007.

Drier, H. S. (2000). 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.

Gigerenzer, G. (1994). Why the distinction between single event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129–161). Chichester, England: Wiley.

Hancock, C. (1995). The medium and the curriculum: Reflections on transparent tools and tacit mathematics. In A. diSessa, C. Hoyles, & R. Noss (Eds.), *Computers and exploratory learning* (pp. 221–241). Berlin: Springer-Verlag.

Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist, 27*(3), 337–364.

Kaput, J., Noss, R., & Hoyles, C. (2002). Developing new notations for a learnable mathematics in the computational era. In L. English (Ed.), *Handbook of international research in mathematics education* (pp. 51–75). Mahwah, NJ: Lawrence Erlbaum.

Khalil, K. I. (2005). *Expert-novice differences: Visual and verbal responses in a two-group comparison task.* Unpublished Masters Thesis. Amherst, MA: University of Massachusetts.

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

Konold, C. (1991). Understanding students' beliefs about probability. In E. von Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 139–156). Holland: Kluwer.

Konold, C. (1994). Teaching probability through modeling real problems. *Mathematics Teacher, 87*, 232–235.

Konold, C. (2007). Designing a data tool for learners. In M. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 267–291). New York: Lawrence Erlbaum Associates.

Konold, C., & Higgins, T. L. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193–215). Reston, VA: National Council of Teachers of Mathematics.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2004). *Data seen through different lenses*. Amherst, MA: University of Massachusetts. http://www.umass.edu/srri/serg/papers/index.html. Accessed 6 Nov 2007.

Konold, C., Kazak, S., Lehrer, R., & Kim, M.-J. (2007). *To understand a distribution, try building it from scratch*. Paper presented at the Fifth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-5), University of Warwick: Warwick, UK.

Konold, C., & Lehrer, R. (in press). Technology and mathematics education: An essay in honor of Jim Kaput. In L. D. English (Ed.), *Handbook of international research in mathematics education* (2nd ed.). New York: Routledge.

Konold, C., & Miller, C. (2004). *TinkerPlots™ dynamic data exploration* (Version 1.0). Emeryville, CA: Key Curriculum Press.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259–289.

Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., & Mayr, S. (2002). Students' use of modal clumps to summarize data. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on the Teaching of Statistics*. Cape Town: International Statistical Institute. CD ROM.

Lecoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics, 23*, 557–568.

Lehrer, R., Konold, C., & Kim, M.-J. (2006). *Constructing data, modeling chance in the middle school.* Paper presented at the Annual Meeting of American Educational Research Association, San Francisco, CA.

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction, 14*(1), 69–108.

Lehrer, R., & Schauble, L. (2007). Contrasting emerging conceptions of distribution in contexts of error and natural variation. In M. Lovett & P. Shah (Eds.), *Carnegie symposium on cognition: Thinking with data* (pp. 149–176). New York: Lawrence Erlbaum Associates.

Pratt, D. (2000). Making sense of the total of two dice. *Journal for Research in Mathematics Education, 31*(5), 602–625.

Russell, S. J. (2006). What does it mean that "5 has a lot"? From the world to data and back. In G. F. Burrill & P. C. Elliott (Eds.), *Thinking and reasoning with data and chance: Sixty-eight yearbook* (pp. 17–29). Reston, VA.: National Council of Teachers of Mathematics.

Saldanha, L., & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*, 257–270.

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making, 10*, 33–51.

Shaffer, D. W., & Kaput, J. J. (1999). Mathematics and virtual culture: An evolutionary perspective on technology and mathematics education. *Educational Studies in Mathematics, 37*(2), 97–119.

Shaughnessy, J. M., & Ciancetta, M. (2002). Students' understanding of variability in a probability environment. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on the Teaching of Statistics*. Cape Town: International Statistical Institute. CD ROM.

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

Wilensky, U. (1997). What is normal anyway? Therapy for epistemological anxiety. *Educational Studies in Mathematics, 33*, 171–202.