LEVERAGING EPISTEMOLOGICAL DIVERSITY THROUGH COMPUTER-BASED ARGUMENTATION IN THE DOMAIN OF PROBABILITY

DOR ABRAHAMSON, MATTHEW BERLAND, BENJAMIN SHAPIRO, JOSHUA UNTERMAN, URI WILENSKY

This article [1] is a case study of technology-facilitated argumentation Several graduate students, the first four authors, present and negotiate complementary interpretations of a diagram generated in a computer-simulated stochastic experiment Individuals use informal visual metaphors, programming, and formal mathematical analysis to ground the diagram, *i.e.*, to achieve a sense of proof, connection, and understanding. The *NetLogo* modeling-and-simulation environment [2] serves to structure our grounding, appropriating, and presenting of a complex mathematical construct We demonstrate individuals' implicitly diverse explanatory mechanisms for a shared experience. We show that this epistemological diversity, sometimes thought to undermine learning experiences, can, given appropriate learning environments and technological fluency, foster deeper understanding of mathematics and science.

Computers can be powerful tools for learning mathematical concepts. One powerful way to use computers for learning mathematics is through the exploration and construction of computer-based mathematical models and simulations (Feurzeig and Roberts, 1999; Jacobson and Kozma, 2000; Wilensky, 1997) However, users' learning experiences through computer-based modeling, we believe, can be greatly amplified beyond running and observing simulations - learners can engage in modes of discourse that challenge the veracity of and assumptions underlying these models and act on these challenges

In order to take greater advantage of the computer medium, we contend, learners should engage in technology-supported argumentation, including questioning the assumptions of existing models and authoring their own simulations. This contention, inspired by Papert's *constructionism* (1991), is developed in this paper through a descriptive and collaborative introspection into a rich and authentic learning experience we shared through critiquing each other's computer-based design work. We attempt to demonstrate the thought processes motivating individuals engaged in creating (programming) mathematics models. [3] In the context of model design, we construe programming not as an end in itself but rather as a natural rhetorical mode of expression that harnesses the computer - the "protean machine," as Papert (1980) calls it - in extending, elaborating, and grounding mental simulation into the public space, a mode that is available to computer-fluent individuals and could be made available to all learners.

Our approach builds upon the literature that advocates that students construct mathematical understandings through engaging in activities within "mathematical environments" (Noss and Hoyles, 1996; Papert, 1991; Piaget, 1952). Computer simulations afford opportunities for such contextualized activity (Feurzeig and Roberts, 1999; Jacobson and Kozma, 2000; Wilensky, 1993, 2001) in collaborative learning environments, where students can 'connect' their qualitative intuitions to formal quantitative articulation such as graphs and formulae.

Collaborative learning differs from exclusively individual learning in that collaboration constitutes a catalyst for argumentative rhetoric, through which individuals articulate hitherto implicit interpretive models (Cobb and Bauersfeld, 1995; Edelson, Pea and Gomez, 1996; Guzdial, Hmelo, Hubscher, Nagel, Newstetter, Puntambekar, Shabo, Turns and Kolodner, 1997; Stahl, 2000). Also, there is great heuristic-didactic value in shifting between different interpretive models for making sense of observed phenomena and between isomorphic mathematical representations such as diagrams, graphs, and equations (Post, Cramer, Behr, Lesh and Harel, 1993). A collaborative phenomenological-mathematical negotiation affords opportunities for formulating and bartering interpretive models (Cobb and Bauersfeld, 1995).

We take the narrative form of first presenting four different interpretations of a simulated probabilistic phenomenon authored in the *NetLogo* modeling and simulation language [2] as part of design research carried out at the *Center for Connected Learning and Computer-Based Modeling* (CCL) at Northwestern University, US. Individual contributors, graduate students in Wilensky's research group in the Learning Sciences and Computer Science departments, explain the experiential grounds for their respective personal interpretations. These personal introspective explanations begin from idiosyncratic constructions of the probabilistic situation, including cogent associations from prior knowledge that these individuals bring to bear in their sense making. Through social interaction revolving around the probabilistic simulation, these individual interpretations feed off each other, converge using shared representations of the mathematical problem, and are woven into an inter-subjective co-constructed account of the phenomenon. Thus, the narrative form of this article is useful, because it conveys an authentic collaborative learning process, giving content to and mirroring the argument we develop. The narrative culminates in a conversation through which we came to see the correctness, the value, and the problematics of each other's points of view. [4]

In the discussion, we collectively reflect on our collaborative learning to argue for the centrality of computer simulation as a vehicle of proof. At the same time, by exposing the disparity between our mathematical assumptions relating to a single representation, we critique the epistemological basis of the ostensible agreement we had achieved.

The mathematical object

Imagine the following computer simulation (see Figure 1). Three "boxes" are set in fixed positions in a row. At the press of a button, another set of three boxes randomly paint themselves either 'green' or 'blue.' Thus, the result of the compound event is, either, green-green-green, green-green-blue, green-blue-green, ..., with a total of 8 different permutations. Now, further imagine that a user creates a "secret key," say 'blue-green-blue,' and then the computer searches for this key. The computer's "unintelligent" search algorithm is to simultaneously paint each of its boxes either green or blue, randomly, and hope for the best. An event, the computer's single guess, can be either a "failure" (key was not matched) or "success" (key was matched). If events are recorded as a list of failures and successes, with each successive event added on at the end of the list, they form a string of length *n*, where *n* is the total number of events (failures + successes), *e.g.*, ffsffffsfsfssffsffffffsfsffsssfsffsfsffs. [5]

These flashing colored boxes are an example of a stochastic experiment, the study of a succession of random events produced by some mechanism, whether concrete or virtual, for the sake of understanding probabilistic aspects of the mechanism. In stochastic experiments, we expect that, if we produce sufficient data, the outcome parameters will reflect some general probabilistic trait of the mechanism (*i.e.*, the Law of Large Numbers, where outcome distributions converge on the probabilities inherent in the mechanism). Computers are useful in rapidly generating data. The computer generates random numbers and these may be instantiated through code as interface objects – *e.g.*, the colors of the boxes – that give sense to the experiment.

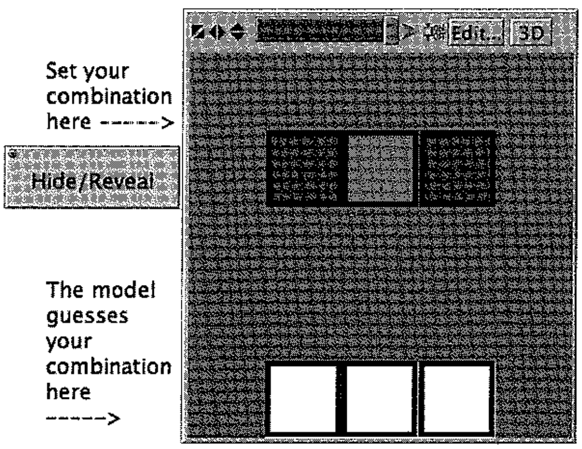


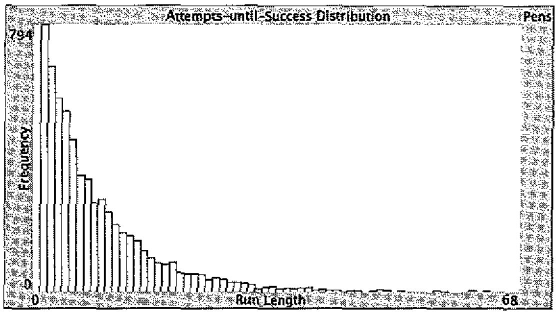
Figure 1: The mathematical object (fragment from a NetLogo *model). The user has set the "secret key" 'blue-green-blue,' and the program will try to guess this key. Monitors and graphs will keep track of hits and misses*

The colored boxes experiment could have been instantiated by flipping three coins instead of coloring three boxes each with one of two colors. Whether showing boxes or flipping coins, the computer-screen interface between the user and the code affords a dynamic perceptual experience that is richer than pressing a button and immediately receiving an output string of events or just a single processed value. Also, the parameters by which experimental data are processed may vary, and, in fact, once data are collected, different experiments may dictate different analyses of the same set of outcomes, that is, different ways of parsing and quantifying the string of failures and successes. For instance, the boxes experiment may be run in order to evaluate the frequency of successes (see Figure 2a), but you might look at the same set of data and wonder about the average number of trials from one success to the next (see Figure 2b). Each analysis can be represented in a different type of graph. As it was, we discussed a graph representing the outcome distribution for number of attempts until success.

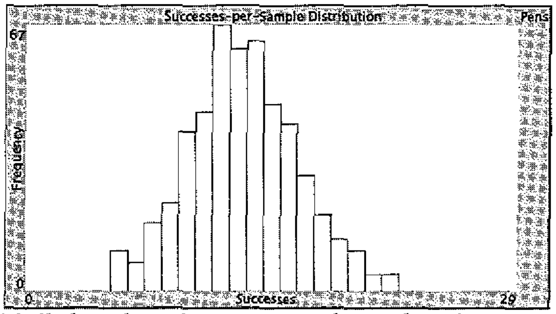
These distinctions between the mathematical constructs and the metaphorical objects and between different models of the same mathematical data as well as issues of how representations inform interpretations of data all usually remain opaque, because learners have no reason to probe their implicit understandings. We had operated under the implicit assumption that once a mathematical construct is instantiated in both a metaphor and a graph that represents the accumulating outcomes of the probabilistic experiment, there would not be much room for individual interpretation - ostensibly, the constructivist ultrasound would not reveal interpersonal differences. But we were wrong. We begin by describing the interaction that instigated the debate, and then we outline what it took for each student to connect to – to *really* understand – the stochastic experiment. The Rashomon structure of the texts enables a conveying of authentic learning experiences of individuals within a collaborative learning space. [6]

Narratives

During a design-team meeting, Dor demonstrated how a computer-simulated stochastic experiment he had authored [7] resulted in a bell-shaped histogram. Dor's approach to revealing the probabilistic traits of the model had been to use sampling. That is, Dor's model parsed the string of individual outcomes into substrings of fixed length, counted up the successes in each of these substrings, and displayed the successes-per-sample as a histogram, which – as it happened – recurrently grew into a bell-shaped curve. The sample size in this experiment was 100 attempts.



(a) A 'ski slope' graph representing the frequency distribution for the number of attempts until success (number of trials until you match the key successfully)



(b) A bell-shaped graph representing the number of successes per sample (number of correct matches per fixed-sized sample)

Figure 2: Two graphs of the same data-set of outcomes in a probability experiment: some of us expected the attempts-until-success graph to be bell-shaped, too, and this expectation provoked the modeling and argumentation reported in this article

Ben, Josh, and Matthew all expressed curiosity during the meeting as to whether or not collecting large samples is necessary for demonstrating the probabilities inherent to the model. In particular, they questioned why one could not simply collect samples of unit-size one (*i.e.*, individual guesses) and count the number of single-guess samples until each successive success. Ben and Matthew were convinced that 'successes-per-sample' would usually mirror 'samples-per-success.' Perhaps the implicit assumption here was that since the search algorithm itself would not be changed and since the variables are held constant - same number of boxes, colors, and total number of attempts - the graphic representation, too, should remain unchanged. Matthew and Ben expected these reciprocal ratios (samples/success → successes/sample) to correspond with simple symmetry transformations of the corresponding distribution curves. Dor explained that he had tried using this attempts-per-success technique and had been frustrated with its results; that the graph produced resembled a ski-slope that had its peak at 'successes on the first guess' and then decreased exponentially as the number of guesses increased.

Josh, Matthew, and Ben all argued that their method was identical to Dor's sampling technique except that their method curtailed each search at the first success to create variably sized samples that contained single successes. [8] That is, instead of taking many samples of fixed size and counting up the varied number of successes in each sample, they suggested counting up a fixed number of successes - 1 - within necessarily variably-sized 'samples.' A graphic representation of the distributed frequency of such sample sizes, they argued, should therefore be identical to the bell-shaped distribution that Dor's technique discovered. Uri recommended that they think about "independence." About this time, the meeting ended.

Dor's world

It's not that Dor didn't understand the graph. He was perfectly happy to believe that the code he had authored himself indeed results in that ski-slope graph: "This is what you get when you run this stochastic experiment." But then his peers, who were witnessing the graph for the first time, challenged it, saying it should be bell-shaped and not shaped like a ski-slope. Perhaps if they saw the graph in a textbook they would not have been so critical, but they were all sufficiently versed in programming to appreciate that Dor may have erred in attempting to formulate computer procedures that emulate the experiment. Spurred by their challenge, Dor had to defend and warrant the graph as a valid representation of his experiment, so he searched for a means to connect to the graph.

Dor typically grounds mathematical constructs in real-world objects and situations (Abrahamson and Wilensky, 2003). So, he struggled to find a situated model that would explain the logic of the '$1/x$-type' curve of the attempts-per-success frequency distribution, and specifically its non-normal shape. [9] Dor came up with the "sticks" model, as follows: imagine that each per-success string of outcomes is a stick of 3, 5, 2, ... units of length, making up a concatenation of sticks with the total length of 40 units: ffs-ffffs-fs-fs-s-ffs-ffffffs-fs-ffs-s-s-fs-ffs-fs-ffs (the same string of data from the *mathematical object* section, above).

Now, the stochastic model that had generated strings of outcomes occurring over time was translated into segments of substantive material extending in space - sticks. From the perspective of the sticks model, the question of the $1/x$-type curve becomes, "Why is it that if we collect sticks of total length N (here, 40) we typically get a greater number of shorter sticks as compared to longer sticks?" If we were looking at this string of f's and s's as one of many different possible outcomes of an event of length 40 total attempts, we could ask, "How many different arrangements of sticks of lengths 1 unit through 40 units are there that sum up to 40 units?" Answering this question mathematically could determine whether or not most collections of addends of 40 do indeed contain more 1-sticks (sticks of length one unit) than 2-sticks (sticks of length two units), more 2-sticks than 3-sticks, more 3-sticks than 4-sticks, and so on. That would explain why there are $1/x$-type curves and not bell-shaped curves on numerous runs of the "green/blue boxes" model. Stripping this down to bare numbers, we are asking the following: "Given an inexorable pool of numbers 1 through 40, how many different arrangements can we form under the condition that each sums up to 40?"

The first observation is that there is just a single arrangement for a single stick of length 40: {40}, a single stick of length 40 units. There are 2 arrangements for a stick of length 39 units: {39 + 1} and {1 + 39}. For a stick of length 38 there are 5 arrangements: {38, 2}, {2, 38}, {38, 1, 1}, {1, 38, 1}, and {1, 1, 38}. And so on. Thus, the shorter the stick, the more different arrangements it may fit into. So, in a random bounded string of total length 40, the shorter the stick, the higher its chance of being included. This explains why the graph descends from 1 through 40. Much later, Dor learned that he had 'discovered' partition-function distributions. [10]

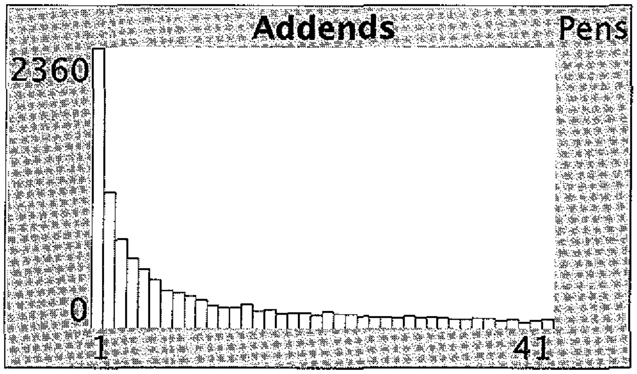


Figure 3: Dor's stick model implemented in NetLogo. *Note its similarity to Figure 2a*

Dor created a *NetLogo* model to simulate his stick gathering so as to have empirical evidence to support the viability of his stick interpretive model of the graph (see Figure 3). He designed the simulation so that it would plot as a histogram the frequencies of each stick over 10,000 runs of the model, in each of which the model randomly selected numbers, adding them up to a specified total (40 in the current example). When we run this model over and over, we receive different specific numbers in the list, but the general frequency distribution, expressed in the histogram shape, remains constant. To all appearances, this is precisely the shape we receive when plotting the attempts-per-success data from a single extended run of the "green/blue boxes," so Dor felt that he now truly understood the graph.

Ben's world

In preparing for the original meeting, Dor had consulted with Ben, who wrote an analysis of the computational complexity of the problem, which Dor then implemented, in the form of a *NetLogo* model (see Figure 4), in his presentation to the research group.

The model examines how the problem's sample space grows with the number of boxes and colors. In performing this analysis, Ben discussed the expected performance of several guessing strategies that one (or one's computer) could use to find a secret key. His analysis relied upon long run averages over large numbers of keys or upon non-random guessing patterns that made guesses non-independent (whereas Dor's scheme's random guessing made each guess's probability of success independent). This work allowed Ben to have a set of strong beliefs about the properties of the model, which were then tested and refined over the course of discussions about the model. However, neither Ben nor anyone else in the group immediately realized that Ben used a strategy wherein guesses were non-independent, whereas Dor's model treated guesses as independent.

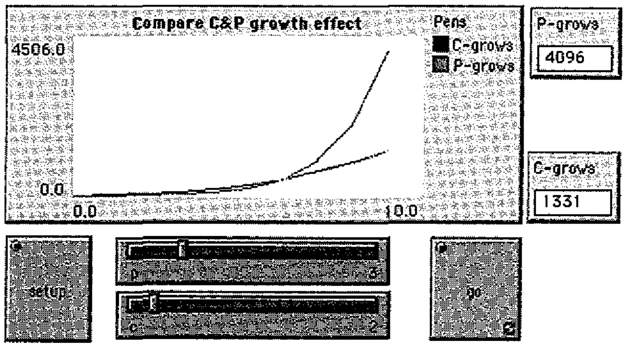


Figure 4: Ben and Dor's representations of search-algorithm expectancies, set to 3 boxes (P) and 2 colors (C)

Matthew's world

Matthew concurred with Ben's analysis of the problem space and embraced this analysis in his own attempt to rethink the search algorithm. While a brute-force key search mapped well onto the problem, a proof of the curved distribution seemed remote. Matthew decided that it is possible to guess randomly in the search space to find a "success," but without any history or pattern to these guesses, the searcher is doomed to repeat "failure" guesses randomly and indefinitely. How could you make an informed guess about the running time of the search through the key-space if it was exponential and memory-less? It would take a very long time to find successes in any large search space.

Ben and Matthew together

Ben and Matthew initially thought that the until-success approach would produce a bell-shaped curve (see Figure 2b). That is, they expected a run of attempts-until-success of length 'mean - 1' to be equally likely as a run of length 'mean + 1.' This sense of balance can seem correct at first, when you reason according to the following logic that conveys a sense of 'equivalence': If you are randomly guessing a number between 1 through x, you are no more likely to guess any of these numbers – they are all equally likely, with a probability of $1/x$. However, the surprising fact is that this line of reasoning does not imply that repeated attempts-until-success will result in an even or a symmetric distribution of guesses. Much of the confusion, we later realized, was embedded in the classic difference between independent and conditional probability – a difference we had all studied yet were not attuned to apply.

Ben and Matthew, dissatisfied with the lack of resolution at the end of the meeting, began writing a *NetLogo* model that implemented their attempts-until-success algorithm. The *NetLogo* model ran according to the following simple algorithm: pick a random number, increment a counter by 1, and if the number is a match, save the counter to a list and reset that counter to 0. They used these data to plot a histogram of the list of samples-until-success counters. Surprisingly, when this code was run, it showed up as the precise graph that Dor had drawn on the whiteboard during the research meeting: a $1/x$-type graph. Ben and Matthew checked the algorithm and the code several times and then formulated preliminary theories to explain the graph. That is, once they were satisfied that they had debugged the code, they reluctantly turned to debug their own thinking - the computer model they had themselves created now constituted an epistemic authority that forced them to reconsider their prior assumptions. They had no clear idea why the graph worked as it did, but now they had some theories. They had ownership of it as a problem instead of as a mistake.

Josh's world

Josh was baffled by Dor's rationale for plotting successes-per-sample. The bell shape of the graph (see Figure 2a) felt correct, but Josh thought it was perhaps unnecessary to resort to sampling in order to get this shape. Specifically, Dor's bell-shaped fixed-sample distribution suggested to Josh that he could represent the attempts-per-success frequencies, too, in terms of a bell-shaped distribution. "Sure," he thought, "it's possible to get to the solution of the key color-combination quicker than the *mean* number of attempts-per-success, but for every time one finds the solution slightly quicker, there'll be a different time when it takes longer - it sort of balances out - just like the normal curve."

Josh proceeded to analyze the probability of a run (attempts until success) that lands in each column of the frequency-distribution graph. That is, Josh attempted to reconstruct the building blocks of the histogram by stepping along column-by-column from the *y*-axis towards the right and accounting mathematically for each step. On each trial (attempt), there would be a 1-in-8 chance of success. That part Josh knew to be true. So, 1 out of every 8 runs should end up in the column representing 1 trial until success, and the rest of the runs - 7 out of every 8 runs - will end up somewhere to the right of that column (see Figure 5). Then, given that a run failed on the first trial, it once again has a 1-in-8 chance of success in the second trial, so $7/8 \times 1/8$ of the total runs will end up with 2 trials until success. This is certainly less than in the first column. Similarly, failure on the second trial would push the run into the next column to the right. This process continues so that, for example, 3 trials until success will happen $7/8 \times 7/8 \times 1/8$ of the time, or $(7/8)^2 \times 1/8$. This implies that the 3rd column should have less than the 2nd column. Josh was convinced this process would continue and that therefore, by induction, the ski-slope was correct after all.

Denouement

When Ben and Matthew came into a subsequent meeting, they were excited. They had coded up a *NetLogo* model to implement the thought process that they, as well as Josh, had had during the earlier meeting and had found Dor's assertion, that plotting trials until success results in a ski-slope graph, to be true. Josh quickly sketched for them his thought process on the problem, and they confirmed that they were then thinking about "something like that." Matthew, Ben, and Josh all, still, expressed dissatisfaction with Dor's stick analogy.

A premise of the stick model is that the probability of a single string of length 1, 2, 3, or more occurring is equal. In other words, translating this framework back to the original problem, after each success it is equally likely that a subsequent string (attempts until success) will run 5 attempts until a success as it will run 2 attempts until a success. However,

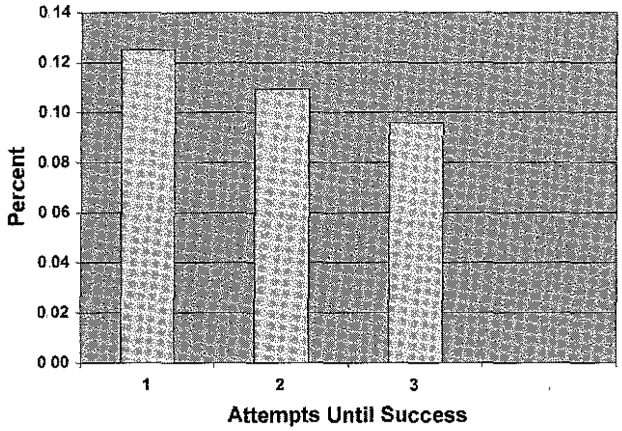


Figure 5: A geometrical progression emerges through considering the chances of achieving a success over exactly 1, 2, 3, and more attempts.

armed with their new understanding of the outcome distribution, where 5 attempts until success has a smaller chance of occurring compared to 2 attempts-until-success, Dor's sticks did not make sense to Matthew, Ben, and Josh. Dor translated this critique in terms of the stick model and realized that he had conflated method and substance - though he had employed a probabilistic solution procedure (*Monte Carlo* brute force), the situational context of the sticks problem is essentially not probabilistic. Also, Dor had confused the phenomenon with superficial properties of its statistical analysis: you cannot take an outcome distribution and make it into its own sample space. Sometimes a stick is just a stick. [11]

Discussion

Mathematics can be a difficult domain for learners. More so, when the subject matter does not fit well with intuitive knowledge, as is often the case with probability (Kahneman and Tversky, 1982; Konold, 1989; Wilensky, 1993, 1995, 1997). Conversely, intuitive knowledge may afford a powerful personal resource for concretizing abstract ideas (Wilensky, 1991) and thus assimilating and appropriating these ideas (Papert, 1980). This tension between, on the one hand, the unintuitiveness of some mathematical ideas and, on the other hand, the value of intuiting mathematics is a polemical, pedagogical, and design challenge that invariably entails tradeoffs. We believe in a mathematics education that goes deeper than merely building isomorphism between equations and other formal representations - the mapping must also be anchored in intuition, that is, assimilated to each individual's collection of models. Connecting to ideas that are counter-intuitive is challenging because you must build mental models that are loyal to both the mathematical constructs and your intuition. That is, you must forge a *middle ground* that reconciles immutable equations and your fickle sense of proof.

As individuals, we had each succumbed to over-lenience in evaluating the validity of our own proofs. We could be so indulgent because, for each of us, the "proof" was not rigorously mathematical, but lay in our personal sense of conviction in the viability of our own models for explaining how the sloped graph came to be. Until the group critiqued our individual convictions, we were complacently entertaining different models for the same graph, because the function of these models was personal and not externalized Moreover, our personal criteria for accepting or rejecting the ski-slope graph were anywhere between vague and unarticulated.

Finally, to varying degrees, we were satisfied to accept the results of a computer simulation. So, only through exposing, sharing, and debating these implicit models could we begin - as individuals and as a group - to critique our underlying assumptions and models. It is perhaps coincidental that as a group we employ a diverse range of explanatory mechanisms for grounding our mathematical understanding - real-world phenomena, programming, and mathematical models - yet this epistemic wealth would have remained untapped and unshared if it were not for our learning environment that fostered argumentation.

Dor's model was essentially mathematically correct, yet proved non-isomorphic to the problem at hand, because it modeled a mathematically *different* phenomenon. Ben and Matthew's models were correct and pertinent to the problem but unintuitive to Dor and Josh. Josh needed mathematical proof to understand a mathematical object. And yet, for each of us, the use of idiosyncratic models as mathematical objects scaffolded learning by providing an *epistemic form* (Collins and Ferguson, 1993) that served in a dialogue both between human and mathematics and between human and human.

All of us held radically different conceptions of what sufficient proof would consist of in this situation (see Figure 6). Dor, coming from a cognitive-psychology background and working primarily in mathematics-education design, was looking for intuitive ways to transform the temporal constituents of the problem (successive stochastic occurrences) into spatial and tangible constituents (the sticks), towards creating a tractable proof-explanation couched in terms of visible objects in the world. Ben and Matthew were looking for assurance that the simulation reflected their set of algorithmic specifications. For Ben and Matthew, it was sufficient for a model produced according to their own specifications to behave identically to a model produced to other specifications to believe that the semantics of the models were identical Josh, being a mathematician, was looking for a formal mathematical proof.

If we were each living and working within a social void, perhaps our individual interpretive models would have sufficed, as inaccurate and/or incomplete as they were. We are all

Author	Strategy	Benefits	Challenges
All	Formal visual metaphors (*e g.*, histograms)	Shared representation of process product	Product over process: shared understanding of phenomenon may mask misunderstanding of underlying process
Dor	Informal visual metaphors	Grounds mathematical object	Not necessarily isomorphic to problem; potentially imprecise
Ben and Matthew	Computer model authoring	Precision; accessible construction	Some programming skills necessary
Josh	Mathematical proof	Precision	Expert construction necessary

Figure 6: Tradeoffs in the authors' mathematical reasoning

relatively well versed in all of the proof techniques used by our peers, yet we each chose to internalize the problem differently. Internalized proof, though, once arraigned and ferreted out to the public domain, must stand the test of peers' rigorous critique. Thus, the pragmatic demand of collaboration in our research team teased the tacit models out of each of us and pitted them against each other until we had reached, as a group, a confluence of our different approaches. This confluence, once internalized, afforded us both greater confidence in the specific content we had discussed and conceptual tools that may inform our future modeling of simulated phenomena - each according to his steadfast style

This narrative could be viewed as a distributed-cognition project. None of us held a complete understanding of the problem independently of each other, our proofs, our models, and the technology that enabled our discourse. The computer-based modeling played a central role in creating this distributed cognition, as it made manifest our respective intuitions without explicitly making the interpretations themselves manifest. By using a concrete, computer-created mathematical model, we could each look at a stable object, interpret it, and inspect our interpretation with the group. In other words, the models served us as a platform to tap and share our previous experiences and ideas. Curiously, the positioning of mathematical knowledge as a perceivable taken-as-shared object was both what sparked the initial conflict and the platform for bartering and negotiating over our phenomenology

Conclusion

Seeing is believing, but believing is an inadequate epistemology of mathematics. There lies a conceptual abyss between being able to run a computer simulation and being able to critique it. This conceptual abyss remains covert when we take mathematical constructs for granted, such as in blindly accepting a computer-generated graph as true. At the same time, making this conceptual abyss explicit, to oneself and to peers affords powerful learning experiences.

We have discussed a case in which several students were fortunate to discover the over-simplifications of their individual understandings of a simulated stochastics experiment. Initially, each student harbored a different conception of the model. These individual conceptions were unarticulated and each constituted a limited and incomplete story of the computer simulation. A breakdown occurred through dialogue that challenged the exclusiveness of each conception and forced the individuals to ground their implicit understanding in mathematical-technological artifacts they each authored - artifacts that exposed each personal construction to interpersonal scrutiny, which was motivated by concern over personal stakes The diversity in explanatory mechanisms and cognitive styles that the group enlisted in analyzing the validity of a shared image created not a fragmented but a robust collective understanding of the mathematical phenomenon underlying the image.

Ultimately, each individual sustained their personal intellectual style, yet we believe that it is such negotiation between competing-cum-complementary styles, a negotiation instantiated in vivid constructions, that engenders individual concretizing of abstract ideas (Noss, Healy and Hoyles, 1997; Papert, 1980; Wilensky, 1991). Whereas we espouse learning environments that respect and foster *epis-*

temological pluralism (Turkle and Papert, 1991), we conjecture that such pluralism that lacks interpersonal critiquing of individual 'makes-sense' feelings may miss on a potentially powerful learning mechanism and even hide personal modeling processes *that are mathematically incorrect* That is, we believe in the educational power of distinguishing between the psychology and epistemology of mathematics (Piaget, 1952; Papert, 2000)

We hope to have demonstrated both affordances and constraints of computer simulation of mathematical phenomena, and specifically the dangers of learning in a computer environment in which models remain at a taken-for-granted iconic level. Moreover, we advocate leveraging conceptual diversity through computer-facilitated argumentation that:

- motivates individuals to effortful mathematical inquiry
- pools together many and varied intellectual resources
- provides opportunities for individuals to build fluency in the domain through argumentation, to use expert vocabulary, and to attempt to negotiate the different explanations
- fosters individual construction of a mature epistemology of science and mathematics that distinguishes between phenomena, models, and forms and content of representation, and
- engenders useful and respectful group discourse between individuals who appreciate the potential strength in diversity.

We conclude that whereas computer simulations can potentially facilitate instructional argumentation, the mathematics-education community should be wary of false agreement between interlocutors that may arise through such ostensible sharing of a representation that does not expose epistemological-mathematical disagreement inherent in the interlocutors' underlying assumptions. A computer simulation is a powerful platform facilitating discourse, but it is only through exposing conflicting assumptions that students can fully avail themselves of the opportunities and promises of collaborative computer-supported learning environments

Notes

[1] Based on a paper presented at The International Conference of the Learning Sciences, 2004, Santa Monica, CA.

[2] Uri Wilensky's software *NetLogo* (1999) can be downloaded from http://ccl northwestern.edu/netlogo, accessed August 31, 2006.

[3] The term programming may connote a certain subclass of so-called "old-style" programming languages and authoring environments not designed for learning or ease of use. Recently, there have been positive developments in authoring environments designed specifically for novices (diSessa, 2000; Hancock [12]; Noss and Hoyles, 1996; Repenning, Ioannidou and Zola, 2000; [2])

[4] In the narrative form, we employ the terms "we' and "us" sometimes to mean the four graduate student "conversers" and sometimes to mean the five authors The context disambiguates the referents

[5] The number of successes has been inflated here relative to the above problem due to the constraints of this textual presentation of a computer simulation. There should be a 1:8 success rate.

[6] Note that the mathematical problem in and of itself is not difficult. We certainly make no claims for having made any original mathematical discoveries Yet, our learning experiences through the problem, we find, do shed light on students' challenges, in general, in studying mathematics.

[7] Dor Abrahamson and Uri Wilensky's software *ProbLab* (2002) can be downloaded from http:// ccl northwestern edu/curriculum/ProbLab/

[8] It appears that the construct of 'sample' can be misleading, perhaps due to prior, not necessarily mathematical, associations, *e g.*, must its size be fixed? Need there be more than one sample?

[9] As it turned out, the term "$1/x$-type curve" was not mathematically accurate. However, this is a signifier we used to gain a common foothold in arguing our interpretations of the graph on the computer screen

[10] The general formula that counts the number of distinct partitions of n is:

$P(n) \sim (\exp[\pi \times \text{sqrt}(2n/3)])/4n \times \text{sqrt } 3$ (Weisstein, 2003)

Note that this formula does not account for the number of different distinct permutations of each of these partitions

[11] Ironically, Dor's sticks model converges to a $1/n$ function

[12] Hancock, C (2003) *Real-time programming and the big ideas of computational literacy*, unpublished doctoral dissertation, MIT, available from http://llk media mit edu/papers/ch-phd pdf, accessed August 31, 2006

References

Abrahamson, D. and Wilensky, U (2003) 'The quest of the bell curve: a constructionist approach to learning statistics through designing computer-based probability experiments', in *Proceedings of the Third Conference of the European Society for Research in Mathematics Education*, Bellaria, Italy, available from http://ermeweb.free fr/CERME3/Groups/TG5/TG5_abrahamson_cerme3.pdf, accessed August 31, 2006.

Cobb, P. and Bauersfeld, H (eds) (1995) *The emergence of mathematical meaning: interaction in classroom cultures*, Hillsdale, NJ, Lawrence Erlbaum Associates

Collins, A and Ferguson, W (1993. 'Epistemic forms and epistemic games: structures and strategies to guide inquiry', *Educational Psychologist* **28**(1), 25-42.

diSessa, A. (2000) *Changing minds. computers. learning and literacy*, Cambridge, MA, MIT Press

Edelson, D., Pea, R. and Gomez, L (1996) 'Constructivism in the collaboratory', in Wilson, B (ed), *Constructivist learning environments: case studies in instructional design*, Englewood Cliffs NJ, Educational Technology Publications, pp. 151-164.

Feurzeig, W and Roberts, N (eds) (1999) *Modeling and simulation in science and mathematics education*, New York, NY, Springer Verlag

Guzdial, M., Hmelo, C., Hubscher, R, Nagel, K., Newstetter, W, Puntambekar, S, Shabo, A, Turns. J and Kolodner,, J (1997) 'Integrating and guiding collaboration: lessons learned in computer-supported collaborative learning research at Georgia Tech', *Proceedings of the Conference for Computer Support for Collaborative Learning*, Toronto, Canada, pp 91-99, available from http://www3 cc.gatech edu/projects/lbd/pdfs/integratingcollab pdf, accessed August 31, 2006.

Jacobson, M. and Kozma, R. (2000) *Innovations in science and mathematics education. advanced designs for technologies of learning*, Mahwah, NJ, Lawrence Erlbaum Associates.

Kahneman, D. and Tversky, A (1982) 'On the study of statistical intuitions', in Kahneman, D, Tversky, A. and Slovic, D (eds), *Judgment under uncertainty: heuristics and biases*, Cambridge, UK, Cambridge University Press.

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

Noss, R, Healy, L and Hoyles, C (1997) 'The construction of mathematical meanings: connecting the visual with the symbolic, *Educational Studies in Mathematics* **33**(2), 203-233

Noss, R and Hoyles, C (1996) *Windows on mathematical meanings: learning cultures and computers*, Dordrecht, The Netherlands, Kluwer Academic Press

Papert, S. (1980) *Mindstorms: children, computers. and powerful ideas*, New York, NY, Basic Books.

Papert, S. (1991) 'Situating constructionism', in Harel, I and Papert, S (eds), *Constructionism*, Norwood, NJ, Ablex Publishing, pp 1-12

Papert, S. (2000) 'What's the big idea? Toward a pedagogy of idea power', *IBM Systems Journal* **39**(3/4), http://www research.ibm com/journal/sj/393/part2/papert html, accessed August 31, 2006

Piaget, J (1952) *The child's conception of number*, London, England, Routledge and Kegan

Post, T., Cramer, K., Behr, M, Lesh, R. and Harel, G (1993) 'Curriculum implications of research on the learning, teaching and assessing of rational number concepts', in Carpenter, T., Fenema, E and Romberd, T (eds), *Rational numbers: an interaction of research*. Hillsdale, NJ, Erlbaum, pp 157-196

[The rest of the references can be found on page 19 (ed.)]