Collaborative Interpretive Argumentation as a Phenomenological-Mathematical Negotiation: A Case of Statistical Analysis of a Computer Simulation of Complex Probability

Dor Abrahamson, Matthew W. Berland, R. Benjamin Shapiro,
Joshua W. Unterman, and Uriel J. Wilensky
Center for Connected Learning and Computer-Based Modeling
http://ccl.northwestern.edu
Northwestern University
abrador@northwestern.edu

> "To swallow something in the hope that it may be wholesome is clearly a commitment, and so is every act of seeing things in one particular way."
>
> Polanyi (1958, p. 363)

> "A full understanding of the power and pitfalls of visual representations is no doubt a long way off. But lack of understanding should not block their use in cases where it is clearly legitimate."
>
> Barwise and Etchemenday (1991, p. 23)

Computers are powerful tools for modeling mathematical concepts, but models are not self-evident phenomena. We wish to query the use of mathematical computer modeling as a form of argumentation and discourse. We presume that mathematical knowledge should be grounded in contextualized activity (Piaget, 1952; Papert; 1991). Computer simulations afford opportunities for such contextualized activity (Wilensky; e.g., 1993) in learning environments where students can 'connect' their qualitative intuitions to formal quantitative articulation, e.g., graphs and formulae. We believe in the advantage of collaboration over exclusively-individual learning as a catalyst of argumentative rhetoric, through which individuals articulate hitherto implicit interpretive models (Haroutunian-Gordon & Tartakoof, 1996; Cobb & Bauersfeld, 1995; Krummheuer, 2000). Also, we see great heuristic-didactic value in shifting between different interpretive models for making sense of observed phenomena, and between isomorphic mathematical representations (diagrams, graphs, and equations; Post, Cramer, Behr, Lesh, & Harel, 1993). A collaborative phenomenological-cum-mathematical negotiation affords opportunities for formulating and bartering interpretive models (Abrahamson, 2002a, 2002b).

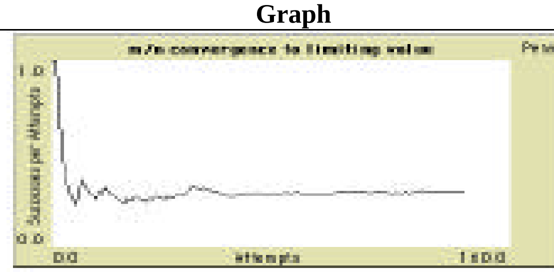
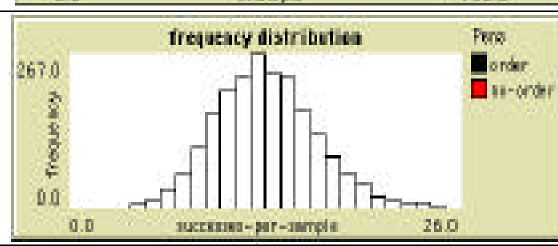
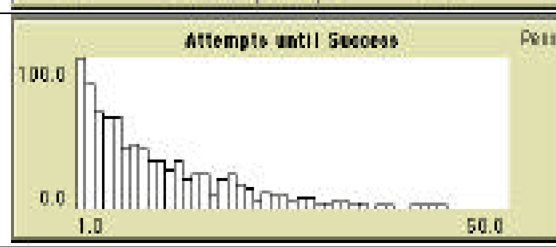
This paper takes the narrative form of first presenting three different interpretations of a simulated probabilistic phenomenon authored in the NetLogo modeling and simulation language (Wilensky, 1999) as part of design research carried out at CCL (The Center for Connected Learning and Computer-Based Modeling) at Northwestern University. Individual contributors—the first four authors, graduate students in the Learning Sciences and Computer Science departments—explain the experiential grounds for their respective personal interpretation. These personal 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 onto a single representation, and are woven into an inter-subjective co-constructed account of the phenomenon. In that, the narrative form of this paper is useful in that it conveys an authentic collaborative learning process, thus giving content to the argument we develop. The narrative culminates in an actual heated debate through which we came to appreciate and problematize each other's points of view. In the discussion, we collectively argue for the centrality of the 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. This critique is related to the broader question of the status and use of visuals in mathematical rhetoric.

The Mathematical Object

Imagine the following computer simulation. Three "boxes" are set in fixed positions in a row. At the press of a button, all 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, etc., with a total of 8 different permutations (2^3 because there are 2 outcomes per box and there are 3 independent boxes). 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 the 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). Whether or not the computer matched the user's key, it records the attempt as either "f" (failure) or "s" (success) and then continues its guessing indefinitely, until stopped by the user. 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[1].

There are several different ways of parsing such an outcome string in order to form statistical assumptions as to the frequency distribution of the 's-' and 'f-' type events and thus inquire into the computer model's probabilistic attributes. Since these outcome strings of f's and s's are the result of an observed phenomenon, and not just generated out of context, it is logical to anticipate that each different parsing is grounded in and motivated by an individual's idiosyncratic cognitive model for interpreting and grouping the virtual simulated events themselves. The variability in interpretive models used for a particular parsing becomes didactically interesting when, in a group of students, each student holds onto a different model. A discussion between these students, each vying for their own model and, possibly, for contradicting statistical conjectures, then becomes a "phenomenological negotiation." That is, because the mathematical inscriptions are coming from a computer simulation, the discussion between these students is not about formulae *per se* but about ways of looking at a phenomenon.

The following three types of graphs of the model's outcomes will be discussed in the paper.

	Unit of Analysis	**Data Processing**	**Graph**
1	Single trial, whether "failure" or "success"	m/n Total # of successes divided by total number of attempts	
2	Samples of a fixed number of trials	Average # of successes per sample	
3	Sub-strings of trials ending with each success	Average length of number of trials until favored event	

[1] 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; alternatively, on can think of this string as a perfectly possible yet highly improbable outcome of this simulation. In the interest of coherency in this paper, we have simplified the model. For further details, and for models mentioned in this paper, see the publications of CCL at http://www.ccl.northwestern.edu . In a public presentation we would run all the simulations we discuss here.

The 1/*x*-type graph representing the third parsing became the focus of the authors' interpretive debate, because each author used a different form of reasoning to 'connect' to the graph, that is, in order to vest the graph within intuition, experience, and knowledge so as to make sense of the graph. This paper questions the didactic adequacy of a 'makes-sense' feeling that is devoid of a critique of underlying mathematical assumptions. We wonder whether a shared 'makes-sense' feeling may not cover a plurality of personal modeling processes. Moreover, whereas we support 'epistemological pluralism' (Turkle & Papert, 1991), we conjecture that the absence of a critique of an inter-personal 'makes-sense' feeling may hide personal modeling processes *that are mathematically incorrect*.

Individual Interpretations

Following, we present the experience of each of the first four authors from their individual perspective. The Rashumon structure enables the conveying of authentic learning experiences. Later, these perspectives are compared and discussed.

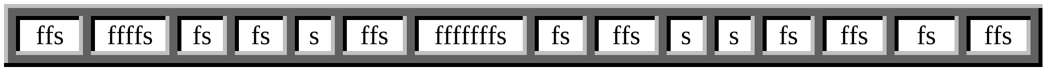
<u>Dor's World</u>:

Bamberger (1991) speaks of students' spontaneous graphic representations of sound sequences as modeling and thus revealing the students' idiosyncratic parsing of the string of auditory stimuli. Likewise, different parsings of probabilistic events reveal different interpretative underpinnings of the meaning of probability. In his protracted self-education in Probability and Statistics through building NetLogo models (see Abrahamson & Wilensky, 2003a), Dor has had opportunities to shift between several perspectives of the meaning of probability. Each perspective can be seen as tightly grounded in a different parsing of the string of f's and s's (ffsffffsfsfssffsffffffsfsffsssfsffsfsffs). It is important to stress that neither of the following interpretations are "correct" or "incorrect." They are each valid in their own way.

1. "f f s f f f f s f s f s s f f s f f f f f f s f s f f s s s f s f f s f s f f s." Taken as a string of 40 independent events, one can sum up the number of successes (15) and compute the probability of a success as the ratio between successes and total outcomes, i.e. 15 / 40 = .375. If the string were long enough, we could argue for the successes-per-events ratio as being the limiting value of this phenomenon. Note that such a perspective entirely ignores the distribution of s's over the string and any variability that could possibly be observed in this distribution.
2. "ffsff ffsfs fssff sffff ffsfs ffsss fsffs fsffs." Taking a statistical perspective, one may parse the string into sub-strings of length 5 events each. Now we can compute the probability of a success occurring in an individual sample: (1 + 2 + 2 + 1 + 2 + 3 + 2 + 2) / 8 = 15 / 8 = 1.875 successes per Sample of length 5 events, or .375 probability of success per single event. Alternatively, first computing probabilities, one ends up with the same value: (.2 + .4 + .4 + .2 + .4 + .6 + .4 + .4) / 8 = .375.
3. "ffs ffffs fs fs s ffs ffffffs fs ffs s s fs ffs fs ffs." In a random string with a total of *n* attempts there is an unknown number of sub-strings, each with a length of 1 through *n* and ending with a success. This interpretative parsing of the events corresponds, perhaps, to an activity in which a success is associated with relief and momentary discontinuity of the search (see Gigerenzer's, 1998, discussion of putative evolutionary underpinnings of probability). Here, the lengths of the strings are 3, 5, 2, 2, 1, 3, 7, 2, 3, 1, 1, 2, 3, 2, 3. Thus, the average length of an attempts-until-success sub-string is 40 / 15 = 2.67. Note that whereas in interpretation #1 we computed 15 / 40 (probability of success per single sample), here we computed 40 / 15 (probability of number of attempts until single success). This reciprocity is no coincidence: In all interpretations, the '40' corresponded to the total number of events and the '15' corresponds to the total number of successes. However, each interpretation harbors a different model of the simulated probabilistic phenomenon, leading to different forms of representation and subsequent statistical inferences.

Interpretation #3 (samples per success) became the bone of contention between Dor and the others, because—counter to their expectation—it does not result in a bell-curved distribution, but rather in a 1/x –type curve. Just as the non-normal distribution of the per-success curve had frustrated and impelled Dor's autodidactic Probability and Statistics (see Abrahamson & Wilensky, 2003a), so it confused the research group and mobilized us to refine, simulate, and argue our personal interpretations towards an ultimate consensus that was larger, we believe, than the collection of individual opinions coming from different perspectives.

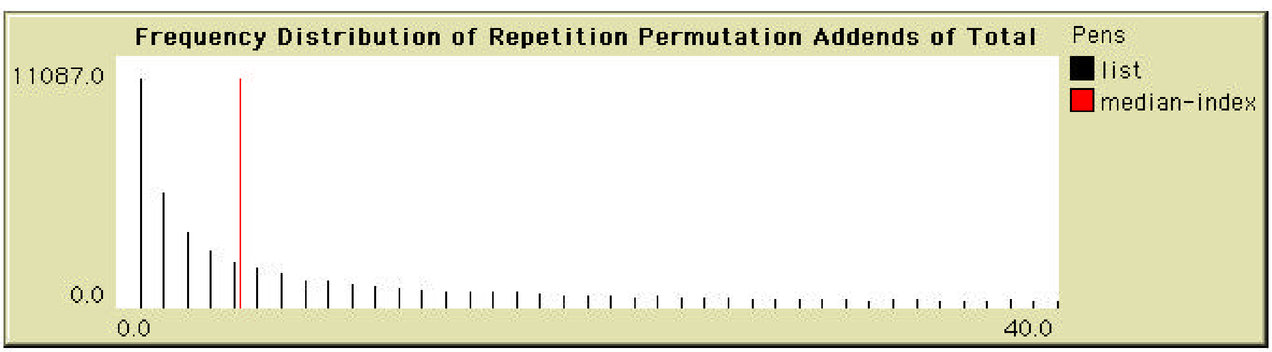
Dor typically construes mathematical knowledge as modeling real-world objects and situations. So he struggled to find an situated model that would explain the logic of the 1/x-type curve of the attempts-per-success frequency distribution, and specifically its non-normal shape. Dor found the “Sticks” model, as follows: Imagine that each per-success string is a stick of 3, 5, 2, etc. units of length, making up a concatenation of sticks with the total length of 40 units:



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 more shorter sticks than longer sticks? If we were looking at this string as one of many different possible outcomes of an event of length 40 attempts, we could ask: How many different repetition permutations of stick lengths 1 thru 40 are there? Answering this question mathematically could determine whether or not most collections of addends of 40 do indeed contain more 1-sticks than 2-sticks, more 2-sticks than 3-sticks, more 3-sticks than 4-sticks, etc. That would explain why one gets1/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 thru 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 stick of length 40: Just that—{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 3 arrangements: {38, 1, 1}; {1, 38, 1}, and {38, 1, 1}. But then this monotonous increase stops being linear, because for a stick of length 37 there are 6 different arrangements: {37, 2, 1}; {37, 1, 2}; {2, 37, 1}; {1, 37, 2}; {1, 2, 37}; {2, 1, 37}. And so on. We thus see that the shorter the stick, the more different outcome arrangements it may fit into. Thus, in a random bounded string of length 40, the shorter the stick, the higher its chance of being included in a concatenation. Moreover, the shorter the stick, the more frequently it is expected to appear in a single random instantiation. This explains the inverse-exponential decrease from 1 through 40 that gives rise to the 1/x-type curve.

Dor created a NetLogo model to simulate his Stick gathering so as to have empirical evidence to support the viability of his Stick model of the graph. 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 added up to a specified total (40 in the current example).



The above histogram represents the following list of frequencies: [10079 5079 3336 2509 1991 1725 1449 1176 1138 994 919 835 777 672 675 657 633 551 533 494 466 434 469 426 405 386 371 345 306 346 308 284 309 316 268 255 269 290 267 244]. The first number in this list, 10079, shows how many 1’s were accumulated, the second number, 5079, show how many 2’s were collected, and so on. The last number, 244, shows how many 40’s were collected. If you multiply the first list-number by 1, the second by 2, the third by 3, and so on until you have multiplied the last number by 40, and then you sum up all these products, you will receive 400,000. This is expected, because we found 10,000 addend-sets of 40. 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.”[2]

---

[2] The vertical red line partitions the total area under the curve into two equal parts. Its height is irrelevant.

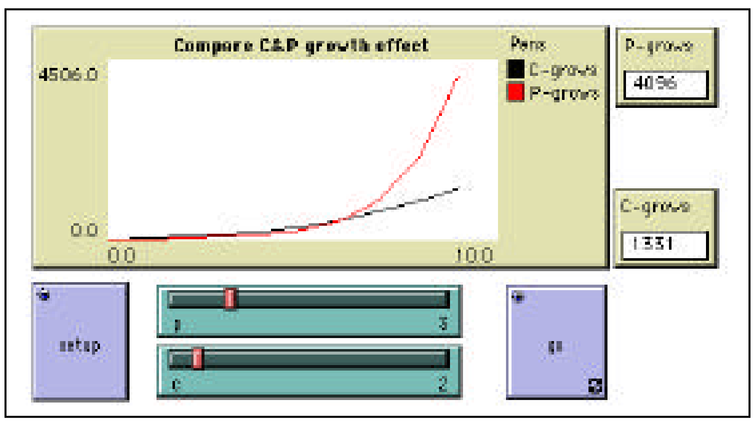
It is interesting to note how using the shape of the graph as the general referent of the phenomenon conveniently eliminates any need to scrutinize the values themselves. The shape is a qualitative re-description of a string of specific integers. The qualities of the histogram lift it above specific lists of integers to the status of 'type,' or 'class' of phenomena (see Stroup, in review). It is this feature of graphs that allows us to speak of bell-shaped or s-curved phenomena. That an icon can facilitate a discussion of classes of phenomena will be pivotal for the inter-personal discourse we will describe in this paper. This discourse flared inadvertently, when Dor was presenting a model:

The colored boxes problem presented in this paper is a simple case of a different model Dor created: The Stochastic Cryptology model. In this model, there can be numerous "boxes" and each can have many colors. Dor showed this model to his colleagues during a research meeting, in which we discussed its prospective inclusion in our design work. One of the extension activities he was suggesting was that students would develop more "intelligent" search strategies for determining the hidden key (color pattern)—algorithms that the students would implement in the code. During the meeting, we ran the model and observed how, for repeated samples of size 100 attempts each, we received a bell-shaped frequency-distribution histogram with a mean of 12.5 successes per sample. This distribution can be reinterpreted as an average of 1 success per 8 attempts or 8 attempts per 1 success.

The Stochastic Cryptology model did not include an "Attempts Per Success" histogram, only the "Successes Per Sample" histogram. Dor had applied an "Attempts Per Success" histogram in a similar model, Equidistant Probability, and, as mentioned above, was disconcerted that this histogram produced a curved downward slope and not a bell-shaped curve. During our meeting, the discussion became interesting when Ben, Matthew, and Josh asked about the graphs. They assumed that an attempts-per-success graph, too, should be bell-shaped. Dor, having "been there," casually drew a curved downward sloping graph on the whiteboard, and stated that if we were to plot the attempts-per-success, we would get such a curve.

<u>Matthew and Ben's world</u>:

After a discussion with Dor about the Stochastic Cryptography model, Ben wrote an analysis of the computational complexity of the problem, which Dor then implemented in his presentation to the research group, in the form of a NetLogo model (see Figure on right). 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. He explained that the size of the problem space grew exponentially with the number of boxes, but polynomially with the number of colors (see graph, on right, where the black line reflects the increase in problem space per increase in number of colors, and the red line—per number of boxes, or "patches"). Ben also described how one could improve the likelihood of finding a key by iteratively running through a circular list of successive keys from a random starting position (i.e. a list where a search starts at a random place and checks successive keys, modulo the size of the keyspace, until success). This could cut the average number of guesses needed across a series of experiments (using different secret keys) from C^B to (C^B)/2 (where 'C' stands for the number of colors, set at 2, and 'B' stands for the number of boxes, set here at 3). The details of this analysis are not as important as the fact that it was done at all, because it allowed Ben to have a set of very strong beliefs about the properties of the model, which were then tested and refined over the course of discussions about the model.



Matthew concurred with Ben's analysis of the problem space and embraced this analysis in his own attempt to coach the search algorithm in terms of the traditional computer-science approach. He tried to apply a computational-complexity model to the mathematical object. While the brute key search mapped well onto the problem, the proof of the necessity of the curved distribution seemed remote. Matthew analyzed that one can 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 one 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. At best, Matthew now had a strong sense that Dor's search algorithm was somehow exponential, but the computer-science perspective *alone* was insufficient for Matthew to implement this sense as a structured proof. In an alternative approach, Matthew attempted to attack the problem using statistics conceptual tools. Matthew's statistics training has been essentially unconnected (Wilensky, 1993), situated in terms of equations and distributions. These conceptual tools—which had served him well on traditional problems—were unhelpful for this untraditional problem. That is to say that his ability to apply statistics outside of traditional problem sets was limited. Thus, neither computer science nor statistics were of much help in initially viewing the problem.

During the meeting, Dor demonstrated how an experiment resulted in a bell-shaped histogram. The sample size in this experiment was 100 attempts. 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 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. Dor explained that he had tried using this attempts-per-success technique and had been frustrated with its results. He told the group that the graph produced resembled a ski-slope that had its peak at success on the first guess and decreased exponentially as the number of guesses increased. Josh, Matthew, and Ben were all surprised, shook their heads, and decided that Dor must have made a mistake. They argued that their method was identical to Dor's original technique except that their method curtailed each search at the first success to create variably sized samples that contained single successes. [3] A graphic representation of the distributed frequency of such sample sizes, they argued, would be identical to the distribution Dor's technique discovered. Uri made a passing remark about "independence." About this time, the meeting ended.

Ben and Matthew initially expected a run of attempts-until-success of length 'mean - 1' to be equally likely as a run of 'mean + 1.' This sense of balance can seem correct at first, when you reason according to the following model 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.

(An additional conceptual challenge inherent in the simulation pertains to the size of the fixed-size samples, and specifically to the relation between the sample size and the possibility space. Because, in the "unintelligent" algorithm for guessing a key within a possibility space of size X, the mean number of attempts-per-success is itself X, unless the sample size is larger than the space of possibilities half of all samples will not contain a success. Such runs yield a distribution graph that peaks at 0, then curves sharply downward. And yet, for simulations involving a sufficient number of boxes and of colors, samples must be so large that pragmatic constraints of running the computer make sampling literally intractable.)

Ben and Matthew, dissatisfied with the lack of resolution at the end of the meeting and convinced that their confusion could be attributed to a mistake on Dor's part, returned to Matthew's office to discuss the problem. They began writing a NetLogo model that implemented their attempts-until-success algorithm (see code and graph on next page).

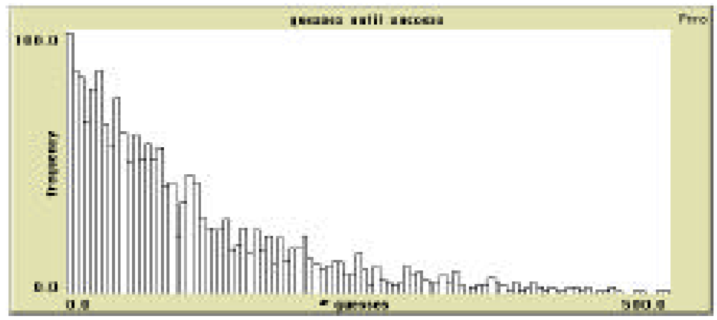
The NetLogo model ran according to the following simple algorithm: Guess a number randomly, check if it's correct, increment a count, and if you're successful, you save the counter to a list and reset it. Also in the code were procedures for plotting a histogram of the list of samples-until-success. Surprisingly, when this code was run, it showed up as a 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. They had no concrete idea why the graph worked as it did, but one theory was that they were looking at some

---

[3] It appears that the construct of 'sample' can be misleading, perhaps due to prior associations, e.g., must its size be fixed? Also, the singular form of the word 'sample' may implicitly support a sense of a sample as a single attempt.

mathematical transformation—possibly a derivative or integral—of the bell-curve. Mostly, they were grasping at straws. However, they were now convinced that this was a problem in itself, something to be proven. They now had ownership of it as a problem instead of as a mistake. Having accomplished their Jigsaw (Brown & Campione, 1996) sub-task, Ben and Matthew, the most computer-science oriented members of the group, decided to approach Josh and Dor with their findings.

```
to go
   set sample-counter sample-counter + 1
   set current-guess random 100
   if current-guess = key
   [   set num-guess-list
         fput sample-counter num-guess-list
       set sample-counter -1
       set current-guess random 100
       histogram-list num-guess-list ]
end
```



<u>Josh's world</u>:
Josh was baffled by Dor's rationale for plotting successes-per-sample. He attempted to couch the problem in terms of attempts-per-success, and yet, the representational forms in Dor's model were obscuring this straightforward task. Specifically, the bell-shaped fixed-sample distribution suggested to Josh that he could model attempt-per-success probabilities, too, in terms of symmetrical relations. Josh was thinking that the plot would be bell-shaped, and that not only would the areas be equal to the left and the right of the mean, but also that the distributions would be the same, i.e., he thought the plot would be symmetrical across the mean. "Sure," he thought, "it's possible to get to the solution quicker than the mean number of attempts-per-sample, but for every time one finds the solution slightly quicker, there'll be a different time when it takes longer. Similarly, for every time one finds the solution *even* quicker, there'll be a different time when it takes *even* longer, etc." There were several problems with his reasoning.

Josh's initial sense of compensation—that within the attempts-per-success distribution probability values could be twinned according to some constant total—was probably based on the notion of probabilities adding up to 1.0. For instance, if we rolled a die with 8 sides, there is a 2/8 chance it will fall on an even number and a 6/8 chance that it will land on an odd number. Similarly, there is a 3/8 chance that it will fall on numbers 1, 2, or 3, and a 5/8 chance that it will not, etc. And yet, Josh was forgetting that there is no guarantee that 8 successive attempts will exhaust the probability space. That is, just because each attempt has a 1-in-8 chance of succeeding does not mean that 8 trials will result in a success. Another problem lay within the statement, "for every time one finds the solution slightly quicker, there'll be a different time when it takes longer."

Dor explained that the graph of trials until success is not, in fact, a bell curve, but is closer to 1/x, a so-called "ski-slope". Josh couldn't make any sense of this, chatted quickly with Ben, and found that Ben agreed. They interrupted Dor and asked for an explanation. Matthew chimed in with Ben and Josh's concern. Given that the meeting was running out of time, Dor answered quickly, but Matthew, Ben, and Josh were not satisfied. Uri responded saying that Dor was correct and also muttered something about independence. Dor told the group that he had found these exact misconceptions in himself and was only able to make sense of them by thinking of his previously mentioned Stick problem. Matthew, Ben, and Josh failed to see the connection. The meeting had already run over, and so the discussion was put aside. The group's confusion was exciting, however, and all exited immersed in the problem.

On his way to their next meeting, Josh made sense of Uri's comment about independence. That a specific trial (attempts-until-success) plotted on the histogram fell in Column 3, for example, meant *ipso facto* that "it did not fall in Column 1 AND it did not fall in Column 2 AND it did fall in Column 3." This dependence, Josh understood, was not represented directly in the attempts-per-success graph and yet understanding this dependence was critical for a true interpretation of the model that the graph represented—an understanding that went beyond recognizing the shape of the graph as representing some class of functions that may or may not be otherwise related.

Josh finally made sense of the problem. He proceeded to find the probability that a run (attempts until success) would land in each column of the frequency-distribution graph. 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 (see table, below). If a run is not successful on the first trial, it cannot make it into the first column—it is "pushed over to the right." This will happen 7 out of 8 times. Given, then, that a run failed on the first trial, there is a 1-in-8 chance of success in the second trial, so "7/8 * 1/8" of the runs will end up with 2 trials until success. Again, failure on this trial pushed the run into the next column to the right. This process continues so that, for example, 3 trials until success will happen (7/8)*(7/8)*(1/8) of the time, or (7/8)^2 * 1/8. Summing up the probabilities of all the columns and pulling the (1/8) out (property of distribution), one is left with a geometric series (with r = (7/8)) so that the total of all the probabilities is (1/8)*(1/(1-(7/8))), which is 1. Josh was convinced that he now understood the problem. {Uri, we will create formulas and equations in a math-editor, ("Tech?") }

Col.	1	2	3	4	5	6	7	8
Prob.	1/8	7/8 * 1/8	(7/8)^2 * 1/8	(7/8)^3 * 1/8	(7/8)^4 * 1/8	(7/8)^5 * 1/8	(7/8)^6 * 1/8	(7/8)^7 * 1/8

When Ben and Matthew came into the subsequent meeting, they were excited. They had coded up a model to implement the thought process that they, as well as Josh, had had during the meeting and 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.

According to the Stick model, the probability of a single string of length 1, 2, 3, or more occurring was considered equal. In other words, according to this framework, 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 steps until a success: A stick is a stick is a stick. However, 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. At least, whereas in Dor's Stick model, 5-sticks were indeed rarer than 2-sticks as addends, the assumption was that these types of sticks were equally ubiquitous on the hypothetical forest floor to begin with, and this assumption of equal ubiquity jarred too strongly with his colleagues' mathematical sentiments.

Discussion

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 in fact the individually necessary sense of conviction in the viability of our model for explaining how the sloped graph came to be. Until the group critiqued our individual convictions, we were complacently entertaining different models, because the function of these models was personal. Sharing these models brought out their differences, both in content and in mathematical-rhetorical impact.

Strategy	Benefits	Deficits
Diagrammatic representations, such as grouping samples, successes, etc.	- Used to intuit strategies for interpreting the mathematical object	- Not necessarily problem-related; - Not necessarily mathematically correct
A computer modeling environment was used to implement the mathematical object.	- More mathematically correct. - Process is visible and intelligible	- Not necessarily the same mathematical phenomenon; - Personal interpretation necessary
Mathematical Proof	- Correct if understood correctly	- Unintuitive; - Expert mathematical interpretation necessary
Histograms	- Shared representation of process product	- Product over process; - Shared understanding of phenomenon may mask misunderstanding of underlying process

Dor's model was the most intuitive but essentially mathematically incorrect. Despite its modeling a mathematically *different* phenomenon, Dor's stick analogy helped him understand the actual mathematical phenomenon being discussed. In this way, he used an incorrect mathematical model to gain insight on a mathematical object. Josh was less satisfied with either the intuitive but incorrect model (Dor) or unintuitive but correct model (Ben, Matthew). 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 that served in a dialogue both between human and math and between human and human.

All of us held radically different conceptions of what sufficient proof would consist of in this situation (see table, previous page). 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 objects in the world. Ben and Matthew were looking for assurance that the simulation was functional to a set of procedural, 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 relatively 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. However, internalized proof, once ferreted out to the public domain, must stand the test of our peers' critique. Strangely, even though we are all relatively well-versed in all of the proof techniques used by our peers, the pragmatic demand of collaboration in our research team teased these tacit models out of each of us and pitted them against each other until we had reached a confluence of our different approaches. This confluence, once internalized, afforded us both greater confidence in the specific content we had discussed and modes of thought that may well inform future modeling of simulated phenomena.

This story 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. Whereas our initial understanding was based on proofs internalized, our eventual understanding was based on proofs expressed. 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, 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 both to tap our previous experiences and ideas and also to look at our own interpretation of the model with others. Curiously, the positioning of mathematical knowledge as a perceivable taken-as-shared object was both what enabled the initial conflict and the platform for bartering and negotiating over our phenomenology. Given this duplicity, are we to trust visualization in mathematics?

The visual in mathematical proof

Over the past century, a certain shift is evident in what scientists are willing to accept as proof. Various authors, whether mathematicians (Davis & Hersh, 1981), logicians, (Barwise & Etchemendy, 1991), and others (e.g., Arnheim, 1969) have advocated foregrounding the visual as a valid rhetorical vehicle. This shift seems to correspond with developments in both analytic (Wittgenstein, 1956) and continental (Heidegger, 1927) philosophy. "[I]ntuition, which logistic had presumably cast out of pure arithmetic and left far behind, has regained its rights (Cassirer, 1950, p. 77). Polanyi (1958) promotes an acceptance of the role of 'tacit knowledge,' primarily visual acumen, as a source of formal thinking:

> I believe that we should acknowledge these sensory actions as proper strivings which we both share and rely on. This endorsement of our native powers of making sense of our experience according to our own standards of rationality should also make it possible for us to acknowledge the ubiquitous contributions made by sense perception to the tacit components of articulate knowledge. And eventually, it should duly condition our manner of acknowledging truth in its articulate forms (p. 98).

This call to promote the visual in reasoning is seen by some postmodern philosophers as a renaissance of sorts. Martin Jay (1988) speaks of vision from the perspective of art history:

> [T]he philosophy it [Baroque vision] favored self-consciously eschewed the model of intellectual clarity expressed in a literal language purified of ambiguity. Instead it recognized the inextricability of rhetoric and vision, which meant that images were signs and that concepts always contained an irreducibly imagistic component. […] [There is a] current imperative to restore rhetoric to its rightful place and accept the irreducible linguistic moment in vision and the equally insistent visual moment in language… (p. 17-19)

The urge to incorporate the 'visual moment' specifically in *mathematical* argumentation is supported by J. R. Brown (1997), who wishes to:

> ...make a case for pictures having a legitimate role to play as evidence and justification, well beyond a heuristic role. In short, pictures can prove things […] Trying to get along without [pictures] would be like trying to do theoretical physics without the benefit of experiments to test conjectures […] Mathematical intuitions are like empirical observations in physics. (p. 161-9)

P. J. Davis (1993), a celebrated mathematician, calls the perceptual vehicle of proof a "visual theorem":

> "[A] visual theorem is the graphical or visual output from a computer program – usually one of a family of such outputs – which the eye organizes into a coherent, identifiable whole and which is able to inspire mathematical questions of a traditional nature or which contributes is some way to ourunderstanding or enrichment of some mathematical or real world situation. (p. 333)

One of several types of visual theorems Davis enumerates are "graphical results of computer programs which the brain organizes coherently in a certain way. Formalizations of this organization and coherence hardly need or hardly can be achieved" (p. 336). Barwise and Etchemendy (1991) contend that "visual forms of representation can be important, not just as heuristic and pedagogic tools, but as legitimate elements of mathematical proofs. […] Visual information can also be integral to the reasoning itself" (p. 9). They advocate "a re-evaluation of the doctrine that diagrams and other forms of visual representation are unwelcome guests in rigorous proofs" (p. 23). Arnheim (1969), too, coming from a background of engineering and architecture, locates an instructional path from pictures to understanding: "Education has to bridge the gap between the bewildering complexity of primary observation and the relative simplicity of that relevant image" (p. 305).

Whilst the image is being invited to guide us from ignorance to enlightenment, still, we are warned by some of the inherent fallacy, the temptation and damnation lurking in falling prey to a meretricious image. Barwise and Etchemendy (1991) alert us to the pitfalls of misperception of an image, and suggest caution in its use. That one can be led astray by an image—that the surest conviction may harbor perceptual distortion—is possibly a function of transparent social forces that lead us to constitute a misleading interpretation of stimuli:

> Between the subject and the world is inserted the entire sum of discourses which make up visuality, that cultural construct, and make visuality different from vision, the notion of unmediated visual experience. Between retina and world is inserted a *screen* of signs, a screen consisting of all the multiple discourse on vision built into the social arena (Bryson,1988, p. 91-2).

Computer-simulated collaborative argumentation that does not scrutinize the image is liable to remain entrapped in its own transparent screen of signs.

Conclusion

R. Hare (1999), in evaluating cognitive explanations of modeling as a scientific tool, says the following: "People, be they scientists or lay folk, do not think by assessing the formal relations between two sets of necessary and sufficient conditions for class membership. Research has shown that concrete prototypes serve the necessary cognitive role" (p. 110). We hope to have demonstrated both the affordances and the constraints of computer simulation of mathematical phenomena, and specifically the dangers of modeling that remains at the iconic level. We conclude that whereas computer simulations afford facilitation of instructional argumentation, the CSCL community should be wary of false agreement between interlocutors that may arise through such a sharing of a representation that does not expose epistemological-mathematical disagreement inherent in the interlocutors' underlying assumptions. A computer simulation is a powerful facilitating platform of discourse, but it is only through exposing conflicting assumptions, for instance as instantiated in different authored code, that learners can fully avail of the opportunities and promises of *collaborative* computer simulations.

<u>References</u>

Abrahamson, D. (2002a). The multiplication table as an ‘object to think with:’ A constructionist project with a phenomenological twist. Unpublished manuscript, Northwestern University, Evanston, IL.

Abrahamson, D. (2002b). “Try to see things my way…while you see it your way”: Examining a tactic for negotiating meaning of manipulatives in teaching-learning argumentation. Unpublished manuscript, Northwestern University, Evanston, IL.

Abrahamson, D. & Wilensky, U. (2003a). *The quest of the bell curve: A constructionist designer's advocacy of learning through designing.* Manuscript under review.

Arnheim, R. (1969). *Visual Thinking*. CA: UC Press.

Bamberger, J. (1991). *The mind behind the musical ear.* Cambridge, MA: Harvard University Press.

Barwise, J., & Etchemendy, J. (1991). Visual information and valid reasoning. In W. Zimmerman & S. Cunningham (Eds.), *Visualization in teaching and learning mathematics*. USA: Mathematical Association of America.

Brown, A.L. & Campione, J.C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (p. 289-325). Mahwah, NJ: Lawrence Erlbaum Associates.

Brown, J. R. (1997). Pictures and proofs. *British Journal for thePhilosophy of Science, 48,* 161-180.

Bryson, N. The gaze in the expanded field. (1988). In H. Foster (Ed.), *Vision and visuality* (pp. 87 -108). New York: The New Press.

Cassirer, E. (1950). *The problem of knowledge: Philosophy, science, and history since Hegel* (W. H. Woglom & C. W. Hendel, Trans.). New Haven: Yale University Press.

Cobb, P., & Bauersfeld, H. (Eds.). (1995). *The Emergence of Mathematical meaning: Interaction in Classroom Cultures.* Hillsdale, NJ: Lawrence Erlbaum.

Davis, P. J. (1993). Visual Theorems. *Educational Studies in Mathematics, 24*, 333-44.

Davis, P. J., & Hersh, R. (1981). *The mathematical experience*. Boston, MA: Birkhauser.

Haroutunian-Gordon, S. & Tartakoff, D. (1996). On the learning of mathematics through conversation. *For the Learning of Mathematics, 16(2)*, 2-10.

Harre, R. (1999). Models and type-hierarchies: Cognitive foundations of iconic thinking. In R. Paton & I. Neilson (Eds.), *Visual representations and interpretations* (pp. 97-111). London, UK: Springer-Verlag.

Heidegger, M. (1962). *Being and time* (J. Macquarrie and E. Robinson, Trans.). NY: Harper & Row. (Original work published 1927)

Jay, M. (1988). Scopic regimes of modernity. In H. Foster (Ed.), *Vision and visuality*, (pp. 3-23). New York: The New Press.

Krummheuer, G. (2000). Mathematics learning in narrative classroom cultures: Studies of argumentation in primary mathematics education. *For the Learning of Mathematics, 20(1),* 22-32.

Papert, S. (1991). Situating constructionism. In I. Harel & S. Papert (Eds.), Constructionism (pp. 1-12). Norwood, NJ: Ablex Publishing.

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

Polanyi, M. (1958). *Personal Knowledge: Towards a post-critical philosophy.* Chicago: The University of Chicago Press

Post, T. R., Cramer, K. A., Behr, M., Lesh, R. & Harel, G. (1993). Curriculum Implications of Research on the Learning, Teaching and Assessing of Rational Number Concepts. In T. P. Carpenter, E. Fenema, & T. Romberd (Eds.), *Rational Numbers: an interaction of research* (pp. 157-96). Hillsdale, NJ: Erlbaum Associates.

Stroup, W. (in review). Understanding qualitative calculus: A structural synthesis of learning research. *International Journal of Computers for Mathematical Learning.*

Turkle, S. & Papert, S. (1991). Epistemological pluralism and the revaluation of the concrete. In I. Harel & S. Papert (Eds.), Constructionism (p. 161-192). Norwood, NJ: Ablex Publishing.

Wilensky, U. J. (1993). Connected Mathematics - Building concrete relationships with mathematical knowledge. Doctoral thesis, M.I.T.

Wilensky, U. (1999). *NetLogo*. Evanston, IL. Center for Connected Learning and Computer Based Modeling, Northwestern University. ccl.northwestern.edu/netlogo

Wittgenstein, L. (1996). *Remarks on the foundations of mathematics* (G.H. von Wright, R. Rhees & G.E.M. Anscombe, Eds.). Cambridge: M.I.T. Press. (Original work published 1956)