STUDENT'S CAUSAL EXPLANATIONS FOR DISTRIBUTION

Theodosia Prodromou[1] and Dave Pratt[2]

[1]Vergina Lyceum, Cyprus; [2]University of London, UK

This paper presents a case study of two students aged 14-15, as they attempt to make sense of distribution, adopting a range of causal meanings for the variation observed in the animated computer display and in the graphs generated by the simulation. The students' activity is analysed through dimensions of complex causality. The results indicate support for our conjecture that carefully designed computer simulations can offer new ways for harnessing causality to facilitate students' meaning-making for variation in distributions of data. In order to bridge the deterministic and the stochastic, the students transfer agency to specially designed active representations of distributional parameters, such as average and speed.

Keywords: causality, agency, stochastic thinking, variation, randomness, probability

VARIATION AND CAUSALITY

This research study builds on ideas which emerged from two research studies: 1) the seminal work of Piaget (1975, translated from original in 1951) and 2) Pratt's work (1998; 2000) as it attempts to clarify how students let go of determinism whilst at the same time re-apply such ideas in new ways to account for variation (Prodromou, 2008; Prodromou & Pratt, 2008).

Piaget and Inhelder (1951) reported how the organism fails in the first place to apply operational thinking to the task of constructing meanings for random mixtures, which were therefore unfathomable. Only much later, according to Piaget, the organism succeeds in inventing probability as a means of operationalising the stochastic. In contrast, students soon gain mastery over the deterministic, appreciating cause and effect at least in a basic manner, apparently lending itself more easily to operational thinking. Instead of interpreting Piaget's work as presenting an impregnable divide between the stochastic and the deterministic, at least until a late stage of development, we began to wonder whether the divide was a manifestation of conventional technologies and whether digital technology might provide a means by which the deterministic might be harnessed to support new ways of thinking about the stochastic.

In Pratt's work (for example, 2000, 2002), students aged 11 years explored computer-based mini-simulations of everyday random generators, such as coins, spinners and dice. These simulations provided functionality beyond that which would be experienced in everyday life. For example, the students were able to change the workings of the simulation and so explore their ways of thinking about randomness. Gradually, the students articulated the heuristic that "the more times you throw the dice, the more even is its pie chart". We detect in this statement a sense that the number of throws *determined* the appearance of the pie chart. Similar causal statements were made about other aspects of the system, such as the effect of changing the workings of the simulation.

Pratt referred to these causal heuristics as *situated abstractions* (Noss and Hoyles, 1996), internal meanings for making sense of phenomena that capture the abstracted nature of the meaning, expressed in language tied to the situation. Pratt and Noss (2002) have further elaborated on the nature of situated abstractions as part of a model for the micro-evolution of mathematical knowledge.

We believe Pratt has made a prima facie case that, in certain conditions, possibly deeply connected to the potential of technologically-based environments, students can construct stochastic meanings out of causality. In this study, we examine this possibility further by building a digital simulation to provide a *window* on students' *thinking-in-change* (Noss & Hoyles, 1996) about average and spread as parameters within a distribution.

First though, we must be more specific about what we mean by causality. In fact, causality can be seen at a variety of levels (Grotzer and Perkins, 2000; Perkins and Grotzer, 2000). Grotzer and Perkins have proposed a taxonomy or a classification scheme that attempts to organise increasing complexity of causal explanation. The taxonomy comprises causal explanations organised in four dimensions along which causal complexity is characterized:

Mechanism includes the most superficial causal explanations, appealing to the most general of phenomena, or to token agents, perhaps "luck", "destiny" or "god's will" in the case of stochastic. Within this dimension we begin also to see inferences of underlying mechanisms.

Interaction pattern begins with simple cause and effect explanations but extends to complex relational causality, involving the co-existence of two or more interdependent factors, possibly with feedback mechanisms. For example, agent A affects agent B but feedback from agent B then affects agent A.

Probabilistic Causality relates to the use of uncertainty in modelling causal relationships. Often apparently deterministic systems hide uncertainty in a chaotic complexity. Thus, does the cup which rests on the table express the equilibrium of underlying static forces? Or should we seek explanation by reference to the chaotic dynamic motion of the sub-atomic particles that constitute the table and the cup? Conversely, we choose to explain phenomena in terms of probability to avoid reference to deep layers of underlying causality. Thus, we might choose to model the outcome from the throw of a dice in terms of probability, rather than by reference to multiple and interacting forces, such as the strength of the throw, the weight of the dice and the friction at the surface.

Agency describes those explanations that recognise that causality is distributed across many elements. Such explanations might use ideas of emergence. For example, we might consider a theoretical distribution as a pattern that emerges from the many pieces of data.

We wished to explore what sorts of computer-based tools might provide us with a window on the use of these differing levels of causal complexity to make sense of distribution, as generated within a computer simulation. We set out to design a virtual environment that supported students in attributing agency to the emergent shape of the distribution while they were discriminating and moving smoothly between data as a series of random outcomes at the micro level, and the shape of distribution as an emergent phenomenon at the macro level.

In that respect, we conjectured that the computer simulation environment could enable students:

- at the micro level to use their understanding of causality whilst at the same time begin to recognise its limitations in explaining local variation, and
- at the macro level to see parameters such as average and spread as causal agents, impacting on the shape of distribution, whilst nevertheless not completely defining the distribution.

METHOD

Approach and tasks. The approach of this research study falls into the design research methodology (Cobb et al., 2003) resulting in the *BasketBall* simulation as depicted below (Fig 1). The animation of the basketball player was controlled by

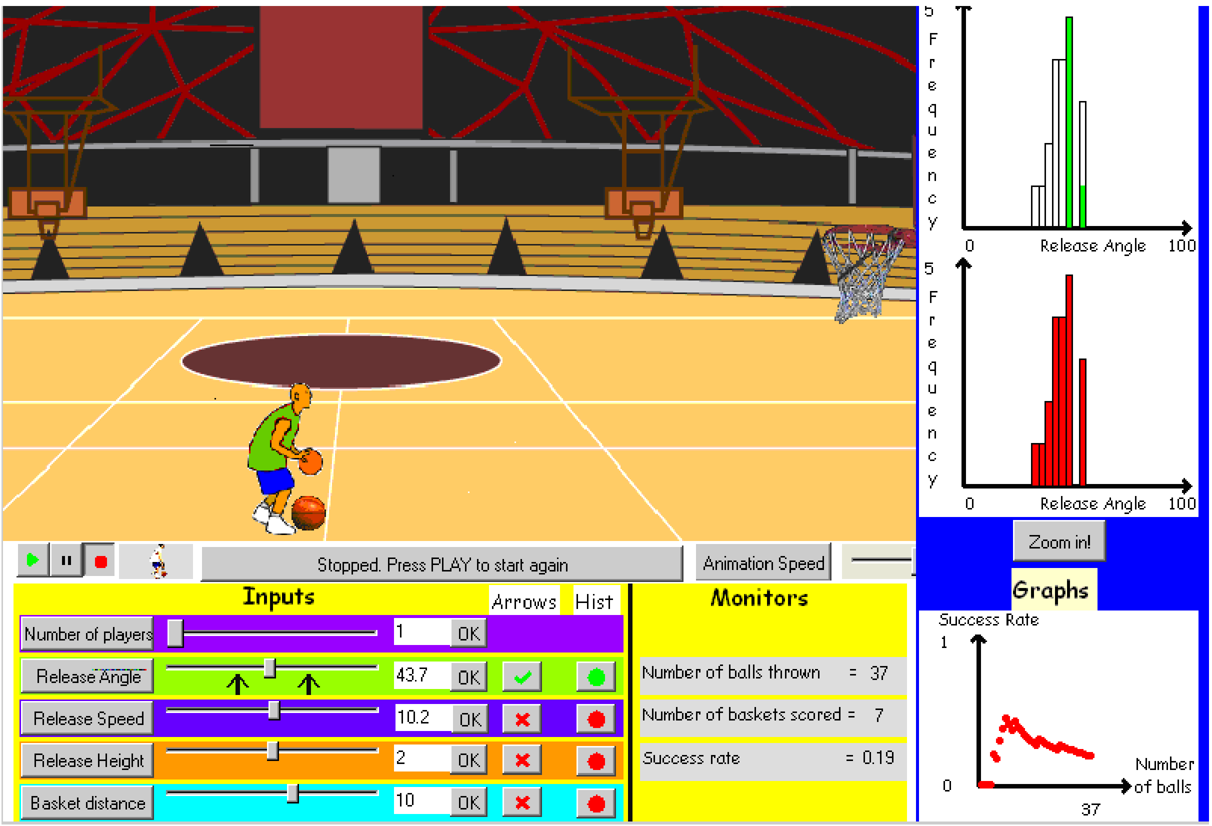

Fig 1: The interface of the *BasketBall* simulation.

varying the handles on the sliders of the release angle, speed, height and distance or by entering the data directly. Once the play button has been pressed, the player continues to throw with the given parameters until the pause or stop button is pressed. The trace of the ball can be switched off. Feedback is made available from the Monitors and Graphs panes. When the arrows button has been switched on, two arrows appear from both sides of the handle on the slider (Fig 2), in which case the value of the parameter is chosen from a distribution of values, centred on the handle of the slider. The students are able to vary these arrows to increase or decrease the spread of the values of the parameter around that centre. The microworld also allowed the students to explore various types of graphs relating the values of the parameters to frequencies and frequencies of success. The students have access to a linegraph of the success rate as well as a histogram of the frequency of successful throws or throws in general against release angle (or release speed, or height, or distance). Initially, the students were challenged to throw successfully the ball into the basket. When the parameters were determined, the histograms of the frequency of successful throws against release angle (or release speed, or height, or distance) appeared as a single bar columns.

Once the preliminary task was completed, some discussion about the realism of the simulation followed, which normally introduced notions such as skill-level, the use of the 'arrows' buttons and the appearance of the histograms. When bias had been introduced to the throws, the graphs appeared as histograms. The subsequent task for the students was to model a real but not perfect basketball player (one who was not successful on *every* throw).

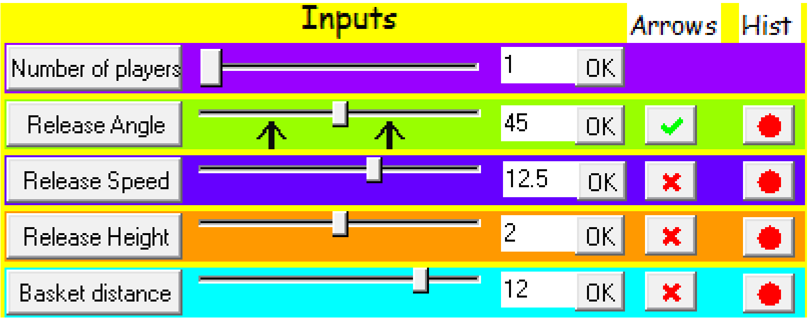

Fig 2: The value of the parameter was selected from a distribution of values, centred on the position of a slider.

Participants. The simulation was used by eight pairs of students in a UK secondary school. It was assumed that the simulation would be used only by students ranging in age from fourteen to fifteen years because a tight focus on the students' intuitions of the distributions indicated that the age of 14-15 years old was mainly ripe for conceptual change in this domain. Another important advantage of working with students of this age was curriculum-based. In the UK National curriculum (DfES, 2000) students of this age are expected to know how to graph data using histograms, dotplots and boxplots, and compare distributions and make inferences, using the shapes of distributions and measures of average and range. Students of this age, therefore, encounter distribution as a collection of data, either given or generated through experiments and surveys.

In this paper, we concentrate on the work carried out by two students, Ethan and Emma (aged 14-15 years), as they engaged with modelling a real but not perfect basketball player. These students had already experienced moving either or both of the arrows, generating values that corresponded to distributions with different spread and bias. The first author was a participant observer during this process. She frequently intervened in order to probe the reasons or intuitions that might lie behind participants' actions.

Data collection and analysis. The data collected included audio recording of the students' voices, video recording of the screen output on the computer, and the first author's[2] field notes. The analysis was one of progressive focussing (Robson, 1993). At the first stage, the recordings were simply transcribed and screenshots were incorporated as necessary to make sense of the transcription. Subsequently, the first author turned the transcript into a plain account. At the third stage, an interpretative account was written by the first author and discussions about the validity of those interpretations with the second author followed, making therefore an account of the data before accounting *for* the activity (Mason, 1994).

FINDINGS

The case of Ethan and Emma provides an illustration of students' typical causal explanations for the observed variation. The two hour session with Ethan and Emma demonstrates how the two students mobilized combinations of different tools to create explanations of variation.

Having already found how to make a successful basket, in the following extract, Ethan and Emma were first introduced to the arrows and they had spent a little time looking at the effect on the animation:

1	Re[1]:	What do you think these arrows do?
2	Et:	…Do they change the angle and the height?
3	Em:	It's just changed the angle, so we will get better results, so we can see.
4	Re:	What do you mean by 'better'?
5	Em:	Because each result is different on the graph (Fig 3).
6	Re:	Why are they better?
7	Em:	Because they much more like realistic.

By looking at the animation, Ethan had recognized that the arrows were causing changes in the throws made by the Basketball player (line 2). Emma refers to the changes in the graph (line 5), and seems to acknowledge that it is more realistic for the basketball player to throw at varying angles (line 7).

A few minutes later however, Emma deliberated upon the role of the arrows in determining the choice of angle:

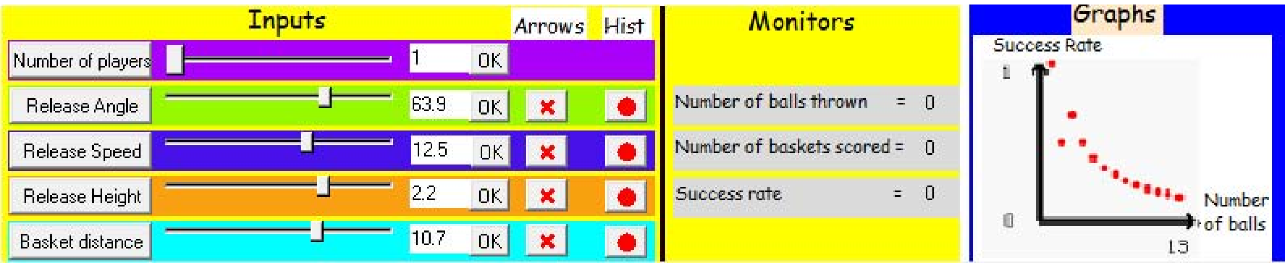

Fig 3: Emma seems to be referring not only to the different values of the angles which were chosen by the basketball player, but also appears to refer to the graph of success rate.

8 Re: What do you think the arrows are for?

9 Em: Is it… where the two arrows are, every time he throws is going to be the distance between that arrow (the arrow to the left of the vertical bar on the slider) and that arrow (the arrow to the right of the vertical bar on the slider)…

10 Re: Do you mean the angle?

11 Em: Yeah … the angle … You can only throw from here to there (pointing to the two arrows). You cannot go any place outside the two arrows.

Emma seemed to be conjecturing that the angle was chosen from between the two arrows (lines 9 and 11), though she still had offered no sense for the mechanism by which the choice was made.

For several minutes, the students experimented with the arrows, at which point their attention was re-focused on the variation which could be perceived through the histograms:

12 Re: Tell me what do you think your graphs will look like. Do you expect these graphs to have one bar, two bars, three bars, or four bars?

13 Em: …about three bars.

14 Re: So, it will not be only one bar? Why?

15 Em: Because he is throwing at different angles… so… he is not throwing at the same angle all the times, so there would be more than one bar.

Emma asserted that variation in the throwing angles would result in additional bars in the histogram (line 15), and soon went further to predict that “the wider apart the arrows around the handle, the more bars there would be in the histogram”. Although, as can be seem, Emma tended to lead the discussion, Ethan was also comfortable at this point that variation could be perceived in the player’s throws and through the frequency histograms.

Their thinking about the relationship between the gap in the arrows and the number of bars was tested further a few minutes later when the bars were moved very far apart:

16 Re: Would there be more or less bars on the histograms?

17 Em: Because he can throw any distance between those two arrows… We haven’t given him a fixed angle to throw it at, so they would not be

the same every time. It will be different… because the arrows give him more of a choice… because the computer like assigns any angle at random between those two arrows… it records it in the graph.

For the first time, Emma referred to a random mechanism operating to make the choice from the gap between the arrows (line 16). She referred also to the interactions between a group of agents (arrows, basketball player, computer), which somehow cooperated to accomplish variation in the distribution.

So far, the discussion had centred on the connection between the gap in the arrows and the variation as seen in the animation or in the graphs. Later, the discussion switched to whether the score was successfully made or not. In the following extract, the handle is positioned on an angle which would successfully throw the ball into the basket and Emma and Ethan know this to be the case. They considered the effect of the arrows on success:

18	Em:	Yeah… because when we put the arrows closer together, so it doesn't have enough choice, like… He can only pick between those two arrows for the release angle… so, he gets a better chance of… to score.
19	Et:	As he's got the release angle inside… that space so… so got to choose that release angle that is scored…
20	Re:	Which is inside …?
21	Em:	63.3… and 76.3… he can only choose… a release angle between those two numbers.

Emma and Ethan both seemed to grasp that a small gap reduced the possibilities for failing to throw a successful basket (lines 17 and 18).

DISCUSSION

As an expert observing Emma and Ethan's activity, it is not difficult to recognise the connection between the arrows and the statistical notion of spread. Such an expert might see the distance between the arrows as a measure of spread. In fact, the data that is actually generated might portray spreads greater or less than that predicted by the gap between the arrows. In this sense the gap between the arrows operationalises the spread parameter of an underlying theoretical distribution, whereas what the students observe is a set of data generated randomly from that distribution.

The above protocol illustrates, through the case of Emma and Ethan, the use of causal explanations, at differing levels of causal complexity, to make sense of variation as it is depicted in the simulated animation of a basketball player and in graphical feedback. These explanations do not take the form of formal robust theory-oriented statements but rather they emerge more as tentative, situated, conjectural utterances, though as the exploration continues the utterances carry more authority and assurance and begin to sound more like conclusions than conjectures.

In Table 1, we list seven observed situated abstractions, based on the body of evidence, which the above protocol typifies:

Ref	Characterisation of situated abstraction	Line
SA1	"the arrows affect the angle that the basketball player throws the ball"	2
SA2	"the arrows affect the graph"	5
SA3	"angles are chosen from between the arrows"	15
SA4	"the wider apart the arrows around the handle, the more bars there would be in the histogram"	15
SA5	"the computer assigns any angle at random between the arrows and records it in the graph"	16
SA6	"The computer assigns a random value from the gap between the arrows for the basketball player to throw the ball"	17
SA7	"the closer together the arrows, the more is that chance to score"	18

Table 1: Examples of situated abstractions

The situated abstraction, SA1, reflects an awareness that the arrows have a causal affect on the variation in throws by the animated basketball player. SA2 similarly recognises a causal effect on the graph. Both these situated abstractions seem to operate at the mechanism level in the Grotzer and Perkins taxonomy. There appears at this stage to be little appreciation of further underlying levels of causal complexity though these begin to emerge later. Situated abstractions, SA3 and SA4, show an increased focus on mechanism as Emma and Ethan strive to make sense of how the arrows affect the player's actions and the appearance of the graphs.

Situated abstraction, SA5, portrays the relationship not as purely deterministic but as including a random element. This introduction of uncertainty seems to represent a move from the mechanism level to probabilistic causality in the terminology of Grotzer and Perkins. Emma and Ethan do not have a sophisticated understanding of probability and so they do not progress deeply into this level but they do seek out, as articulated in both SA5 and SA6, explanations that accept a probabilistic language as a means of coping with a possible multitude of unknown factors. Of course, this move may have been all the easier to make because randomness is something they perhaps regularly experience on computers through, for example, playing computer games.In SA7, Emma and Ethan recognise, even with their ongoing probabilistic language, combinations of agents, as predicted in the interaction pattern level in the Grotzer and Perkins taxonomy. Emma and Ethan envisage a transference of agency from the computer to the arrows and then to the Basketball player. We note that we have previously reported a similar transference of agency from the student itself to the arrows (Prodromou, 2008; Prodromou & Pratt, 2006).

CONCLUSION

The facility to transfer agency seems to be a crucial move in making connections between the causal and the stochastic (from our perspective on the student's psychological state) and in harnessing the deterministic (from the perspective of designing for the student's abstraction). Indeed, by providing handles, arrows and a basketball player, together with feedback on "their actions" (and here we intentionally give these things agency), we set up the possibility that distribution might be seen as generated by the agents. Technological tools, therefore, may have been especially significant in supporting the construction of stochastic meanings out of causality and that in this sense they may provide a route towards operationalising the stochastic in the absence of formal operations.

We believe that such a view of distribution is consistent with the expert position in which a theoretical distribution is sometimes viewed as a generative model, for example sending out a signal *determined* by the average parameter and noise *determined* by the spread parameter. Such a position accepts that the deterministic view of distribution is useful within limitations. Simulations such as basketball might provide opportunities for students to begin to appreciate that expert position.

Even though we have referred regularly to agents, the reader may have noticed that nothing has actually been said about the final level in the Grotzer and Perkins taxonomy, that of agency in which causality is distributed across many agents. In fact, we intend to report elsewhere on students' attempts to make connections from the distribution of data to the theoretical distribution, a direction which demanded an emergent perspective from the students.

When students view variation as an accomplishment of a combination of agents, they think about distribution in terms of a relational model. Their expressions move along the underlying causality dimension towards considering that the simulated *BasketBall* is a context perturbed by a random mechanism. Students' accounts began gradually to address dimensions of probabilistic causality, such as noisy systems, chancy systems. Students were able to view the activities in the *BasketBall* context as noisy processes dependent on a variety of intervening variables. Those accounts were themselves preceded by students' understanding of mediating causality, where predominant causal agents, such as the arrows, and neglected agents of lower saliency in the context, such as the basketball player and the computer, mediate the effect of one agent to another in order to cause variation in the setting (Interaction pattern).

NOTES

1. 'Re' refers to the first named author (Dr. Theodosia Prodromou).

2. The data were collected for the first author's doctoral thesis.

REFERENCES

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9-13.

DfES. (2000). *Mathematics: The National Curriculum for England*. London: Department for Education and Employment.

Grotzer, T. A., & Perkins, D. N. (2000, April). A taxonomy of causal models: The conceptual leaps between models and students' reflections on them. *Paper presented at the annual conference of the National Association for Research in Science Teaching*, New Orleans, LA.

Mason, J. (1994). Researching from the inside in mathematics education - locating an I - You relationship. In J. P. da Ponte & J. F. Matos (Eds.), *Proceedings of the Eighteenth International Conference for the Psychology of Mathematics Education* (Vol. 1, pp. 176-194), University of Lisbon, Portugal.

Noss, R., & Hoyles, C. (1996). *Windows on mathematical meanings: learning cultures and computers*. Dordrecht: Kluwer.

Perkins, N. D., & Grotzer, A. T (2000). Models and moves: Focusing on dimensions of causal complexity to achieve deeper scientific understanding. *Paper presented at the annual conference of the National Association for Research in Science Teaching*, New Orleans, LA.

Piaget, J., & Inhelder, B. (1975). *The Origin of the Idea of Chance in Children* (L. J. Leake & P. D. Burell & H. D. Fischbein, Trans.). London: Routledge & Kegan Paul. (Original work published in 1951)

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

Pratt, D. & Noss, R. (2002). The micro-evolution of mathematical knowledge: The case of randomness, *Journal of the Learning Sciences*, 11(4), 453-488.

Prodromou, T. (2008). The role of causality in the Connecting thinking about distribution. *Unpublished PhD Thesis*, University of Warwick, UK.

Prodromou, T., & Pratt, D. (2006). The role of causality in the Co-ordination of the two perspectives on distribution within a virtual simulation. *Statistics Education Research Journal*, 5(2), 69-88.

Prodromou, T., & Pratt, D. (2008). The emergence of stochastic causality. *Proceedings of the Thirty Second Conference of the International Group for the Psychology of Mathematics Education*, (Vol. 1, pp. 137-144) Morelia, Mexico.

Robson, C. (1993). *Real World Research*. Oxford: Blackwell.