

Students' Construction of Meanings about the Co-ordination of the Two Epistemological Perspectives on Distribution

Theodosia Prodromou¹

¹ School of Education, University of New England, Australia

Correspondence: Theodosia Prodromou, School of Education, University of New England, Australia. E-mail: tprodrom@une.edu.au

Received: March 31, 2012 Accepted: May 11, 2012 Online Published: October 29, 2012

doi:10.5539/ijsp.v1n2p283 URL: <http://dx.doi.org/10.5539/ijsp.v1n2p283>

Abstract

The importance of the connection between theoretical frequencies and observed relative frequencies in pedagogy is advocated by various probabilistic researchers. This study examines an associated area of importance to mathematical education. Little is known about the process by which the co-ordination of the data-centric and modelling perspectives on distribution might be achieved. The focus of this paper is on variation in students' (aged 14 to 15 years) evolving meanings about the co-ordination of two distinct epistemological perspectives on distribution. Extracts from two case studies illustrate students' construction of two interpretations for the two perspectives on distribution through their attempts to transform, directly or indirectly, the specific modelling distribution, and observing how that changes a graph/histogram of the actual outcomes. This is done by using on-screen control mechanisms to change the way that the computer generates the data within a carefully designed computer simulation.

Keywords: data-centric perspective, modelling perspective, distribution, data, variation, randomness, probability, computer simulation

1. Introduction

1.1 Two Perspectives on Distribution

Statistics is gradually becoming a mainstream area in the school curriculum, starting in the secondary grades and continuing through into university. In an introductory statistics course at the university level, the topics of observed frequency distributions and probability distributions are essential (Cohen & Chechile, 1997). In the UK National Curriculum (DfES, 2000), students at the lower and upper secondary levels are expected to know how to graph data using histograms, dotplots, and boxplots, and to compare distributions in order to make informal inferences about underlying populations using the shapes of distributions and measures of range and average. Higher-achieving students should be able to extend their understanding of data distributions to other measures of spread and to frequency density.

In the assessment regime of the National Curriculum, the expectations for graphing ability focus on distributions of data, thus emphasizing a data-centric perspective in the sense that the primary focus is on data—an aggregated set of actual outputs—from which a pattern or model may or may not be discerned.

The data-centric perspective is supported by the introduction of digital technology into schools and a corresponding interest in Exploratory Data Analysis (Tukey, 1977) and innovative software (such as Fathom and TinkerPlots) as a means of engaging students in statistical analysis without using classical probability. International interest in EDA as a method for teaching statistics that places emphasis upon a data-centric perspective is limited, as EDA has not yet offered a coherent statement about how students might abstract from that perspective a rich concept of distribution.

This data-centric perspective stands in contrast to a modelling perspective, in which we attribute probabilities to a range of possible outcomes in the sample space. In the modelling approach, a model gives rise to the variation in data. In the modelling perspective, probabilities are associated with a set of possible outcomes. This perspective regards data as generated by factors such as signal and noise (Konold & Polatsek, 2004), where the signal is the average value and variation the noise around it. Data distributions are seen as variations from the ideal model,

resulting from noise or error randomly affecting the signal or main effect, as reflected in the model itself.

Prodromou and Pratt (2006) argue that the modelling and data-centric perspectives on distribution offer different interpretations of the role of data. In the data-centric perspective, data are the starting point, spread across a range of values, from which a trend or model might be discovered. Prodromou and Pratt refer to that spread of data as portraying variation. In the modelling perspective, data are the outcomes, generated by random selection from the values expressed in the model. Prodromou and Pratt refer to the spread of values in the modelling distribution as portraying randomness.

In order to co-ordinate these two perspectives, it is necessary to see them as a duality that encompasses both the deterministic and the stochastic (Prodromou & Pratt, 2006, p. 72). This paper tries to gain a deeper appreciation of the process by which the co-ordination of the data-centric and modelling perspectives might be achieved.

1.2 Co-ordination of the Two Perspectives

A critical element of coordinating the two perspectives on distribution is constructing meanings for probability and relative frequency (sometimes referred to as empirical probability). Probability is seen as the means for quantifying randomness (the possibility of an occurrence of a random event) within the modelling perspective.

Relative frequencies, which, after repeated random events, stabilize around certain values, play a similar role for quantifying variation in the data-centric perspective.

In the research literature, researchers of probabilistic cognition point out the importance of connecting the empirical probability and relative frequency (Abrahamson, 2008; Chaput, Girard, & Henry, 2011).

Abrahamson and Wilensky (2007) have considered the problem of co-ordinating two or more epistemologies, a process they refer to as “bridging”. To study this, they developed bridging tools, hybrid representations bearing structural properties of both perspectives. The bridging tool is designed to allow students to see both perspectives, thus encouraging connections to be made between them.

One example of their work of particular relevance to this study was a device they called the 9-Block, a 3-by-3 matrix in which each of the nine squares could be either green or blue. The 9-block can be read in terms of combinations (i.e., the number of squares of each colour), or permutations (distinct, ordered arrangements of coloured squares). The study had two main activities:

Activity 1. Using paper, crayons, and glue, students created collections of 9-blocks (avoiding duplicates) and were guided to arrange these in the form of a histogram in which the columns correspond to the different combinations (i.e., 1g8b, 2g7b, ...), thus building a combinatorial sample space of all possible permutations.

Activity 2. Using 9-Blocks modelled in the NetLogo (Note 1) software suite, the students ran a simulation that produced random permutations that were recorded in a histogram of combinations, which grew incrementally, taking a shape similar to the combinations tower (Abrahamson & Wilensky, 2007, p. 12).

In terms of this paper, Activity 1, which considered all possible outcomes (both permutations and combinations), foregrounds the modelling perspective, while Activity 2 foregrounds the data-centric perspective.

The researchers asked students to address the fundamental principle of probability that the shape of a distribution that emerges from randomly generated outcomes gradually becomes similar to the shape of the anticipated distribution produced through theoretical analysis. The students tried to explain why those two representations should be alike in shape. In particular, one student perceived the combinations tower as a theoretical-probability tool that represents propensity and not merely as the collection of all combinations and permutations. So, according to Abrahamson and Wilensky (2007), the combinations tower constitutes a bridging tool between theoretical and empirical probability, creating the idea of a sample space without explicit attention to ‘distribution’ or the idea of a sample space.

Abrahamson and Wilensky (2007) provide a clear example of how students can engage with the modelling perspective by paying attention to the logical combinations that make up the combinations tower. They also report on how the students were able to observe the emergence of the histogram of results, which is a graphic representation of the data-centric perspective.

Their approach to bridging tools conveys the plausibility of designing in order to bring into close proximity epistemological perspectives that might appear distinct to the naïve learner but coherent to the expert. The researcher’s aspiration was not only to design a bridging tool for the two perspectives on distribution, but also to execute a

careful study of how that fusion of epistemologies might take place in practice. The purpose of this paper is not to report on the design of the BasketBall Simulation (see Prodromou & Pratt, 2006, for the process of the iterative design of the simulation), but rather to examine the students' use of the simulation, particularly their thought processes related to the co-ordination of the two perspectives on distribution.

2. Method

2.1 Approach

This study is part of a *designed experiment* (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003), whereby we gain insights about students' thinking-in-change by observing their interactions with a design, in this case the *BasketBall* simulation (Figure 1). In this article, the researcher reports on students' interaction with the fourth iteration of the *BasketBall* simulation (for information about previous iterations, see Prodromou, 2008).

2.2 BasketBall Simulation

The design of the *BasketBall* simulation relied on the notion of phenomenolising (Pratt, 1998), the process of turning mathematical ideas into *quasi-concrete objects* (Papert, 1996) that can be manipulated on-screen by the students, who are provided with the opportunity to make sense of a mathematical concept by using it.

In the process of phenomenolising, the modelling and data-centric perspectives on distribution, randomness, and variation were "controlled" directly or indirectly through parameters that were instantiated as on-screen sliders representing measures of average (the handle on a slider) and spread (arrows bracketing the handle on the slider; see Figure 2). This blending of a measure (the variable) and a representation (the slider on the interface) with a control (the handle on the slider) promotes using before knowing (Papert, 1996).

The *BasketBall* simulation is a simple computer simulation of a basketball player shooting a basketball whose flight is controlled using a simple ballistic simulation. The human user can control several variables that determine the flight of the ball (e.g., angle and speed of release, distance from basket, and height of the release point). The user can also control whether these variables are fixed or whether they vary randomly. When the variables are set to undergo random variation, the variation is either normally distributed and the user can control the variance and the spread of the variation, or the variation is determined by user-defined probabilities.

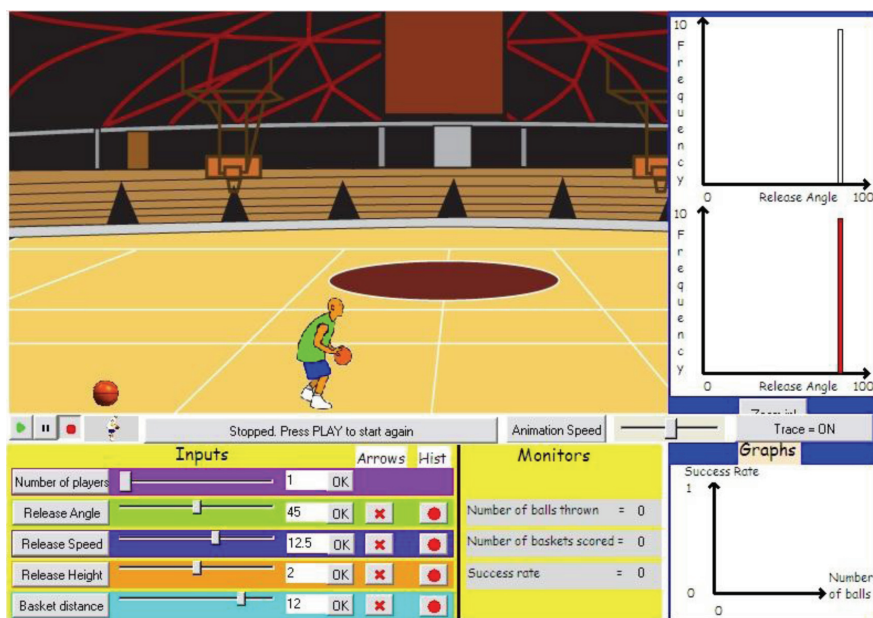


Figure 1. The sliders allow the student to control the simulation variables, including the release angle, release speed, release height, and basket distance. The arrows button controls whether the variable is fixed or varies randomly. Results are displayed in graphs of data, such as number of balls thrown, number of goals scored, and success ratio. In this figure the graph of success ratio against time appears in the Interface window. Other graphs can be chosen through an options menu

The simulation variables can be set by dragging the handles on the sliders (see bottom left hand corner of Figure 1) or by entering a numerical value directly in the box to the right of the slider.

Once the play button is pressed, the computerized basketball player starts throwing the ball with the selected parameters and continues using those parameters until the Pause or Stop button is pressed.

Random variation of the shot-control variables can be introduced by pressing the Arrows button to the right of each slider. In the interface, two arrows then appear on the slider, one arrow to the left of the slider handle and one to the right (Figure 2), indicating the spread of a Normal distribution of values, centred on the position of the slider handle.

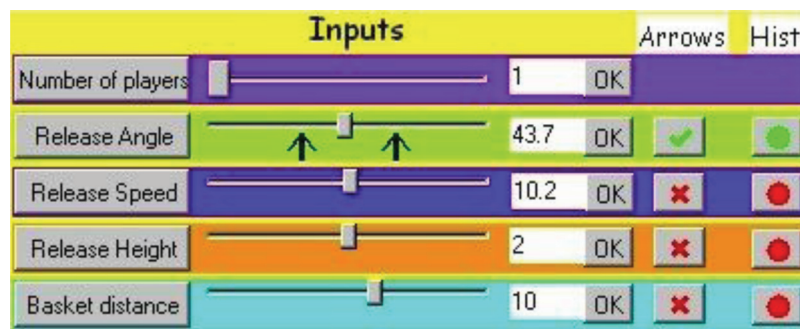


Figure 2. When the Arrows button is turned on, the value of the parameter is chosen randomly by the computer

Either or both of these arrows can be moved, generating values that correspond to distributions with different spreads and bias.

The students have access to a line graph of the success rate of throwing balls into the basket in the bottom right hand corner of Figure 1. The simulation also allows the students to explore various types of graphs relating the values of the parameters to frequencies of attempts and frequencies of success (Figure 3). This display represented the *data-centric perspective* of distribution.

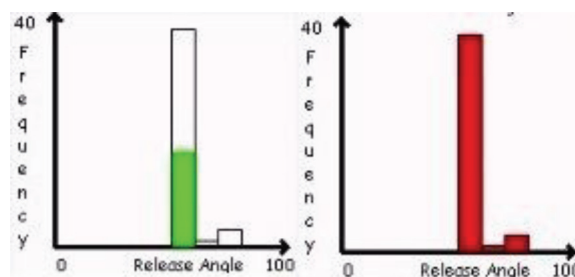


Figure 3. The histogram displays the frequency of release angles used in the basketball throws (right) and also how often these were successful throws (left). Similar graphs could be displayed for any of the parameters

The students also had access to a graphical representation of the modelling distribution (Figure 4). A dialogue box showed the distribution of values from which the computer would randomly choose, given the interface settings chosen by the student.

The students were able to move either the arrows or the handle on the slider and observe the impact of their actions on the graphical representation of the modelling perspective on distribution.

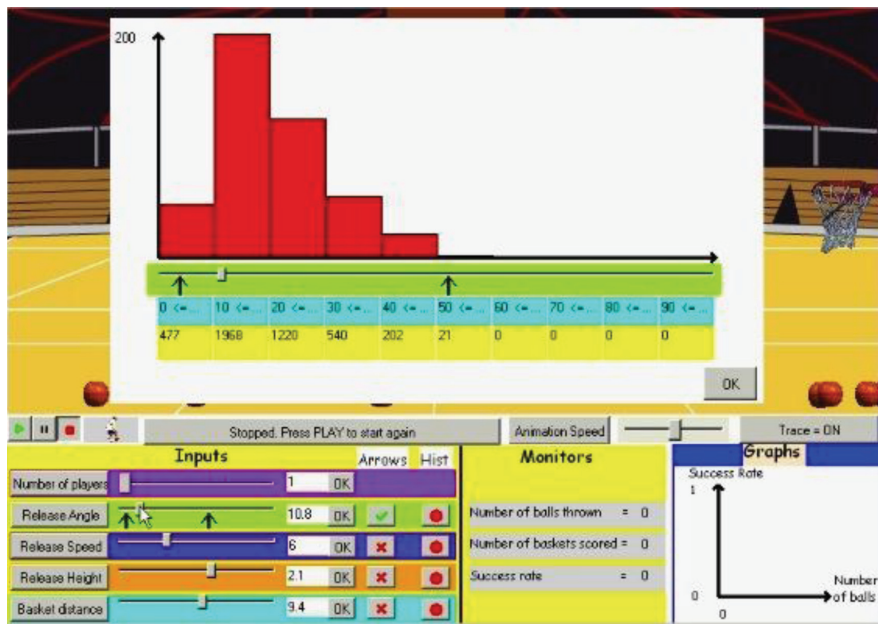


Figure 4. The modelling distribution is shown on the main screen. The students could change the distribution by moving either the arrows or the handle on the slider, or by directly entering numbers in the yellow boxes beneath the modelling distribution

While the simulation was playing, the students had access to both the modelling and the data-centric perspectives on distribution (Figure 5).

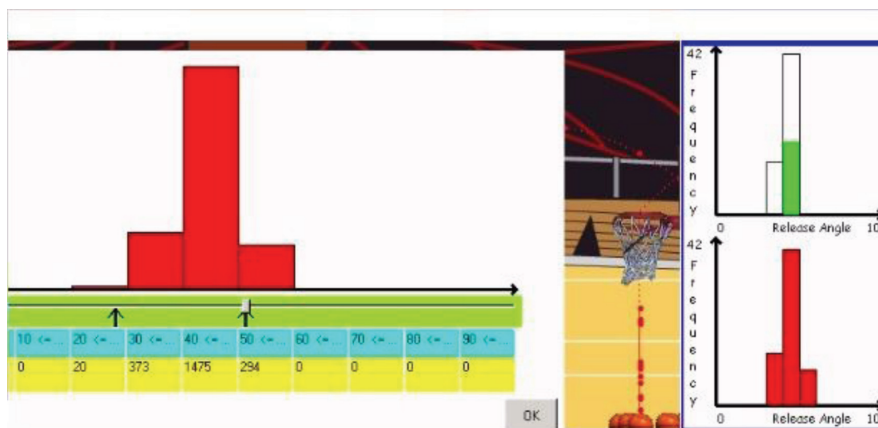


Figure 5. On the left is seen the modelling distribution, showing the way in which the values of the corresponding variable were chosen. On the right, showing the data-centric perspective on distribution, plot the data generated by the computer, showing both successful throws (top) and attempts (bottom)

The students were able to control the model that generated the data, and through this control they had access to information about the more general modelling perspective on distribution. The simulation, therefore, allowed students the facility to transform the modelling perspective on distribution directly, changing the way that the computer generates the data, which allowed indirect control over the data-centric perspective.

2.3 Task Design

Participants were given written instructions for their activities. Students' attention was drawn to the specific interface controls so they would construct situated abstractions (Noss & Hoyle, 1996), or learning heuristics, during

their interactions with the quasi-concrete objects of the Basketball simulation.

Task 1. The students were challenged to throw the ball into the basket by adjusting the parameters that control the throw (release speed, angle, etc.) without selecting the Arrows button that would introduce variation. In this case, if the students found a combination of variables that would lead to a successful shot, every following shot would replay the same path, and thus the histogram would show a single bar.

Discussions with students in the context of the task introduced notions such as skill level, accuracy, and luck (Prodromou & Pratt, 2006).

The primary aim of task 1 was to familiarize the students with the basic controls of the simulation, while tasks 2 and 3 were intended to expose them to different kinds of distributions.

Task 2. The students introduced variation to the throw of the basketball by pressing the Arrows button; they were also able to increase or decrease the spread of the parameter values (Figure 2).

In this task, the histograms that were generated revealed variation, as the parameter values were being selected randomly from a distribution of values set by positioning the handle and the arrows.

Task 3. Students were instructed to open a dialogue box that showed the distribution of values for a variable from which the computer would randomly choose, given the settings of the handle of the slider and the arrows.

The students were able to move the arrows and the slider handle and observe corresponding changes in the graphical representation of the modelling distribution. After a few minutes of simulation play, students were asked to observe the impact of manipulating the arrows and the handle of the slider on the outcome histograms. The students compared the graph of the modelling perspective (visible on the left of the screen), to the graph that plots the data generated by the computer (i.e., the data-centric perspective, on the right of the screen), and attempted to make interpretations of the relationship between the perspectives.

Task 4. The students were invited to manipulate the modelling distribution directly by inputting numerical values for each possible outcome for a given variable (which were then translated by the simulation into probabilities). The simulation, therefore, allowed them another way to alter the modelling distribution directly and the data-centric distribution indirectly.

In designing these tasks, the researcher conjectured that simultaneous access to both perspectives on distribution would enable more systematic exploration of the perceived relationship between the two. During the tasks, the researcher probed students' reasoning about basic features of the distributions such as centre and spread, shape, and skewness. As the students tried to identify elements in the meanings of the two perspectives on distribution and link real data with the modelling perspective of distribution, their discussions were supplemented by the researcher's interventions and questions. These interventions aimed to probe the students' reasons or intuitions and investigate different components of their meanings and understandings.

The students spent varying amounts of time observing the two distributions and reasoning about their different aspects. The researcher's agenda here was to try to afford opportunities for observing expressions of their meanings about the different features of the observed (data-centric) distribution and the model (the modelling distribution), and how well the model manifested as a set of outcomes.

2.4 Participants

The main study consisted of eight pairs of 14 to 15 year old students in a UK secondary school. The selection was curriculum-based, because students of this age will have only encountered distribution as a collection of data generated from an experiment.

Each pair consisted of one boy and one girl from the same class, so interpersonal relationships had already been established prior to the research. The teachers were asked to choose articulate students who would have no difficulty setting up a "friendly" group, and who, in their assessment, were of a "middle" attainment in mathematics.

Students of average ability were chosen on the assumption that they would not construct the knowledge too quickly, allowing more opportunities for observing the construction of new knowledge about distribution.

In this paper, the researcher reports on the interactions of two pairs of students with the simulation while performing task 4. These pairs were chosen as the best examples of students who created the dual connection, and thus their discussion of the formation of this connection was the most useful for analysis; their responses are not presumed to be representative of all the students.

2.5 Procedure

The students worked with the researcher, who acted as a participant observer. The researcher frequently intervened in order to tease out the motivation for particular actions when the motivations were not transparent. Sometimes, when the researcher intuited that the student had more to say, she used several probes, such as asking questions intended to trigger reflection by the students.

Each session was systematically recorded by Camtasia software (TechSmith Corporation). Discussions between students and the researcher were audio recorded. Every screen-based action was captured on videotape. The video recorded the screen output on the computer, but, for ethical reasons, did not record the faces or bodies of the students. The researcher responded to this necessary loss of data by taking brief notes when students' body language or facial expression appeared to be indicative of conceptual evolution. During the exercises, the researcher prompted students to use the mouse to point to objects on the screen when they reasoned about computer-based phenomena during their attempt to make sense of particular phenomena that they could see but couldn't yet fully grasp. This helped supplement the audio recording of the students and clarified their actions and interactions with the simulation.

The recordings of the eight pairs of students were fully transcribed shortly after the end of each session.

Plain case accounts were developed for each pair of students. Those case accounts narrated stories that reported in detail the verbal exchanges between the students and the researcher. They also described complex dynamic and unfolding occurrences of events, students' relationships, and interactions with the software and other factors in the circumstances surrounding the exchanges. The case accounts avoided, however, as far as possible, interpretations of the transcript. This approach was informed by Mason's (1994) notion of giving an account of before accounting for the activity.

Based on these plain case accounts, interpretative case analyses were developed that criticise why and how students' constructions of meanings developed. The case analyses used the transcripts of the pre-interviews as well as the case accounts of the clinical interviews.

The case analyses became the main focus for subsequent analysis and triggered further phases of progressive focusing (Robson, 1993) to identify key foci for ensuing study. Important similarities and differences between the interpretative case analyses were then identified by constant comparisons (Glaser, 1978) of these eight interpretative case analyses. This paper reports on one emergent key focus: the different ways in which the students coordinated the modelling and the data-centric perspectives on distribution.

3. Findings

The story of how the meaning-making unfolds is best told through the detailed analysis of the two pairs of students' interactions with the graphical representations of the two perspectives on distribution, which illustrates how the students made an intuitive synthesis of the two perspectives. The material presented provides the clearest illustration of how these pairs of students co-ordinated meanings that embrace both the data-centric and the modelling perspectives on distribution.

The students used three primary interpretations while performing task 3: (a) general intentionality, (b) stochastic intentionality, and (c) target (see Prodromou, 2008). The following section discusses data and conclusions from students' activity when performing task 4. During task 4, however, the students used only the second and third of the previously mentioned interpretations.

3.1 Creating Their Own Modelling Distribution

During task 3, the students drew conclusions about the relation of the two perspectives on distribution. To follow this, in task 4, they were invited to create their own modelling distributions by defining the possible outcomes and the likelihood of each possible outcome.

The first pair discussed, Anna and James, supplemented an earlier intention interpretation with a target interpretation, while the second pair, Nick and Sarah, introduced an intention interpretation to supplement their earlier target interpretation.

3.2 Anna and James: Task 4

While performing task 3, James noticed that the shapes of the two graphs were becoming the same. James seemed to associate the two distributions by looking closely at the effect on the data-centric distribution of having the

simulated player take more shots. Both Anna and James referred to the modelling distribution as what was intended and to the data distribution as what actually happened. James talked comfortably about variation. Anna and James, however, did not talk about randomness and probability. They were using a sense of a general intention (I_G).

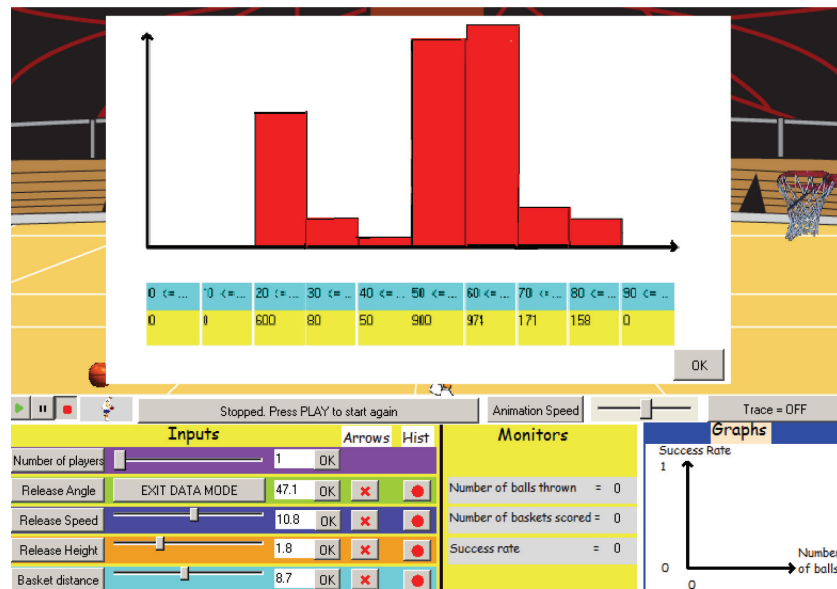


Figure 6. The students changed the way in which the values of the corresponding variable were chosen from the modelling perspective on distribution

Anna and James spent 27 minutes in task 3; they saw the modelling distribution as the intended outcome for the data-centric distribution without, however, specific reference to any random mechanism for selection. They then moved on to the fourth task in which they were able to directly change the values in the modelling distribution (Figure 6). After they had let the simulation play for about 3 minutes, they looked at the data-centric perspective on distribution (Figure 7). At this point their responses started to reveal the material we will report here:

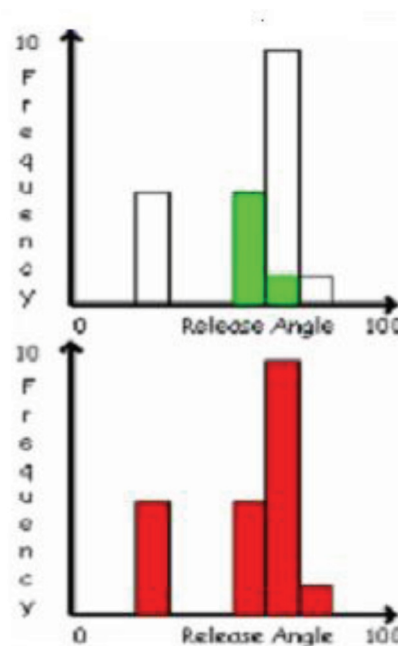


Figure 7. The data-centric distribution

1. James: We haven't used the arrows. Oh, no! Well, yeah on these ones ... the rest of the three, because it's only selected one option ... from the variety we told him he can use one thing, but at the end they've got more, because they've got choice to do that ... Because we did not ask him to select any arrow ... Oh! No ... no ... I don't know.

James was apparently confused because he recognized that the data-centric perspective revealed variation, even though the arrows, which had introduced variation in the earlier tasks, were no longer acting as agents of variation. He attempted to explain the occurring variation in terms of the arrows since he was still relating variation to changes that the arrows had generated. Notwithstanding, it is noteworthy to point out that James talked about “choice to do that”, “options,” and “selections”. Are we to say that the word “option” constitutes an intuition of probability, thus contradicting the researcher's earlier claim about the lack of probabilistic language during their activities with the *BasketBall* simulation? Here one could assume that there could be a probabilistic interpretation since the word “option” almost bridges both interpretations. It is possible, however, that the word “option” has nothing to do with notions of chance or probability, but is simply expressing the idea of an alternative.

The absence of any kind of control over the generational powers of distribution through instantiations of average and spread and arrows caused apparent discomfort to James (line 1). The researcher attempted to place emphasis on students taking on the role of the arrows by editing the modelling distribution and thus transforming it directly. In order to get the students to attend the indirect impact of their actions on the data-centric perspective on distribution, the researcher asked them to compare the graph on the right (bottom) with the graph on the left.

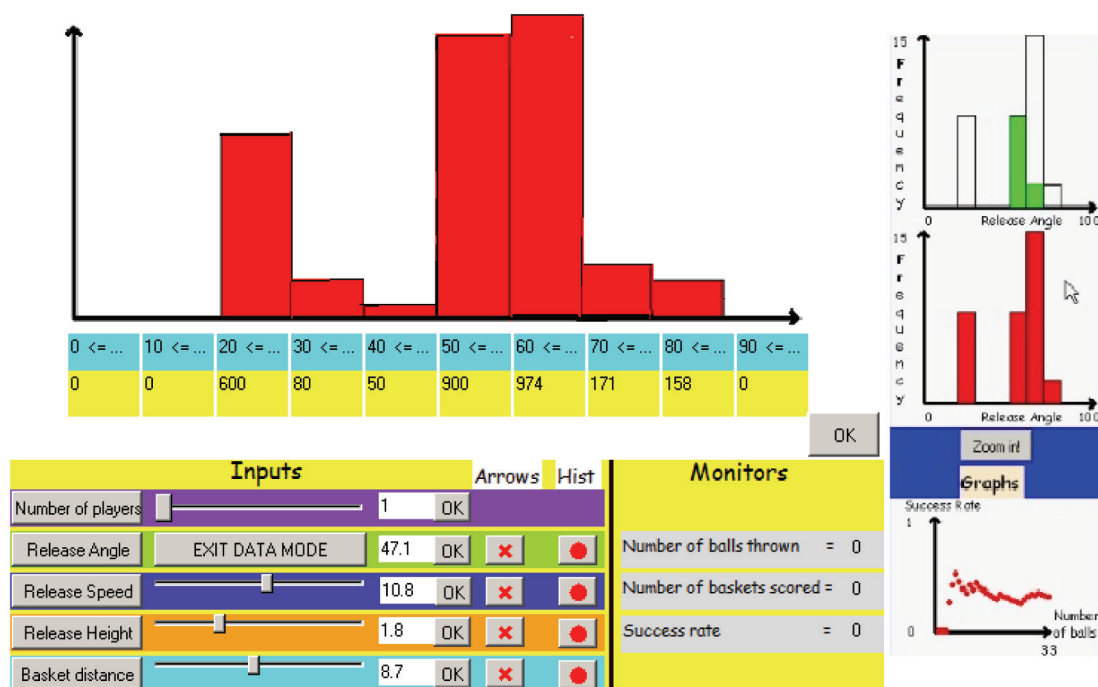


Figure 8. The two graphical representations of distributions

2. James: The last three on this graph (talking about the graph on the right) are the same as the last three on that graph (graph on the left).
3. Res: Yeah, but the other bars are not there.
4. James: I don't know.

James's interest exhibited a tendency to gravitate explicitly towards local reasoning. He paid attention solely to the three bars of the data distribution (50 <=, 60 <=, 70 <=), which looked the same as the corresponding bars in the modelling perspective.

5. Res: So, you said earlier that these three bars (pointing to the 50 ≤, 60 ≤, 70 ≤ bars of the graph on the left) are the same as these three bars (pointing to the 50 ≤, 60 ≤, 70 ≤ bars of the graph on the right).
6. Anna: Yeah.
7. Res: But the 0 ≤, 10 ≤, 30 ≤, 40 ≤, 80 ≤, and 90 ≤ bars are missing.
8. James: Because that (modelling perspective) is just showing you what ... what selection we've made and that (data-centric distribution) is showing which one he is using ... he is using that, just he hasn't used them, because he hasn't selected them to be used.
9. Res: Do you mean that in the future he will select to throw at these angles?
10. James: Yeah ...

To explain why some of the bars were missing from the data distribution, James articulated a one-directional connection of the two perspectives on distribution, perceiving again the modelling distribution as the intended outcome and the data distribution as the actual outcome. This one-directional connection is particularly clear in the following interchanges:

11. Res: So, what do you expect to happen in the long run, after many throws?
12. James: The graphs show all the selections we should want to.
13. Res: Are you are saying that the graph on the right is going to be the same as the graph on the left?
14. James: Yeah.
15. Res: Do you agree with James?
16. Anna: Not ... that level ... Not, all selections.
17. Res: What do you mean by "not all selections"?
18. Anna: Like, quite a few selections but not a lot.
19. Res: Will the two graphs be the same or not?
20. Anna: Not the same, but similar ...
21. Res: Similar? When you said similar, could you please be more specific please?
22. Anna: Like ... some like ... you know we said that the last three bars are the same ... will be the same.
23. Res: So, don't you expect the 30-40, 40-50 bars?
24. Anna: No...
25. Res: What do you think?
26. James: Yeah, I think after a long time they're going to be around, because there are options the tallest to do, so it's likely more ...

Anna disagreed with James. James seems now to have adopted a target interpretation of the modelling distribution (as opposed to earlier, when he was adopting an intention interpretation). However, Anna refers to some bars of the data-centric distribution being mirrored in the modelling distributions, but she does not believe that all of the data-centric distribution would be exactly the same as the modelling distribution. So, while James seems to have constructed a target interpretation (lines 12, 26), Anna has perhaps only constructed a partial target interpretation.

3.3 Nick and Sarah: Task 4

During task 3, Nick and Sarah observed both perspectives on distribution, coordinating information given by the modelling distribution to predict the shape of the data-centric distribution. The students never articulated a clear probabilistic appreciation of how the throws were generated. Nevertheless, they were able to express the notion of the modelling distribution as a target towards which the data-centric distribution is directed.

In task 4, they were asked to directly set the modelling distribution by typing in their own values (Figure 9). The researcher expected that by defining the modelling distribution themselves, the students would, in general, develop meanings of randomness in terms of the way the throws were chosen. In fact, as can be seen in the next excerpt from the session, those on-screen inputs indirectly orient students' predictions (lines 27-33).

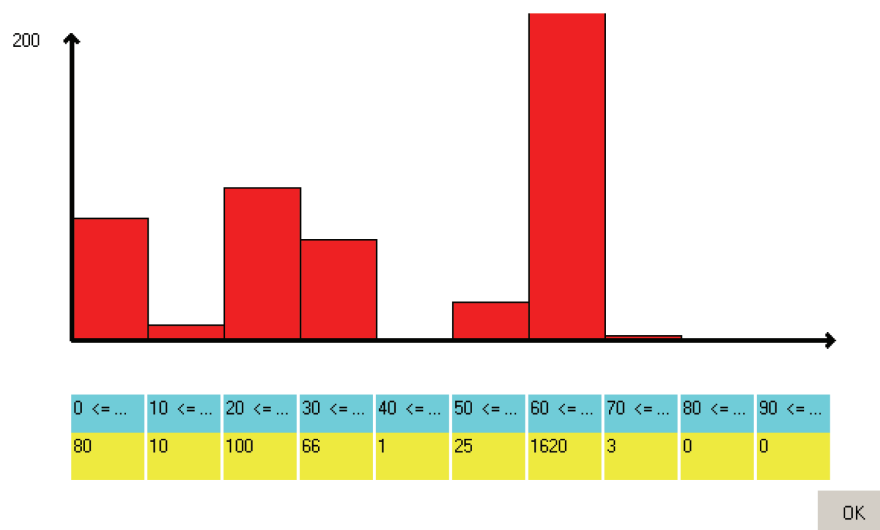


Figure 9. The modelling distribution after students typed in their own values in the intervals

After a few minutes, the researcher suggested that they look at the data-centric distribution (Figure 10).

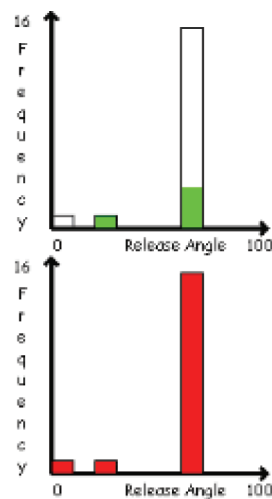


Figure 10. The data-centric distribution

27. Nick: After we've taken 18 shots ... here because it's taken from 2 different points ... after a certain amount of shots ... there would be a bar here (pointing to the 30 <= interval of the graph on the right) ... that would be equal to the one that we did up there (referring to the graph on the left)...

28. Nick: You know, when we clicked on this one (referring to the interval 10 <=), it would be the same ...

29. Res: Why?

30. Sarah: Because we changed (referring to the interval 10 <=) it to 10.

31. Nick: I think ... from 10 degrees to 20 degrees (referring to 10 <=) equals 10, and there would be a bar there ... I think it will happen. If it doesn't shoot at that angle ... the graph up here ... it's got to be 10 of that shot (referring to the interval 10 <=) ... it's only doing three different shots.

32. Res: Are you expecting to have shots from there (throwing at 30 <=, 40 <=, 50 <= angles)?

33. Nick: Yeah ... he should ... He is not taking there yet, that's the difference.

The occurrence of isolated bars (local reasoning) attracted Nick's attention. We see in the above interchange that Nick was somehow surprised because different attempts were chosen only from three intervals.

The researcher discerned that the students lacked an accurate understanding of the probabilistic mechanism by which the throws were selected randomly from the intervals of the modelling distribution. The researcher tried to investigate if the selection could be understood as being due to chance and how the software could support that kind of understanding.

The researcher asked them if they thought that the bars were appearing with an order. They responded as follows:

34. Sarah: It just does it randomly.

35. Nick: Because and this comes (pointing to the 60 \leq bar of the graph on the left hand and right hand) and that came (pointing to the 10 \leq , 30 \leq bars).

36. Sarah: That was there, the bigger one was already there ...

37. Nick: I don't think there is an order ...

The facility of being able to observe the way by which the throws were chosen from the modelling distribution to generate the data-centric distribution appears to have enabled them to observe the order with which the bars were appearing (line 35, 36).

The students' interest was attracted by the occurrence of individual bars and they predicted which bars would appear next and reasoned why certain bars would appear first (local reasoning). After 28 throws:

38. Nick: I think it does it randomly ... wherever it takes a shot ... and then appears on the graph (pointing to the graph on the right hand side) ... wherever it takes a shot from a certain point, then it appears on the graph ... there ... it's only taken one shot from there... (from 40 \leq) since we started ... yes ... He's taken 22 shots from there (pointing to the 60 \leq bar of the graph on the right), so, that's why it is so big (talking about the tallest bar of the graph on the right hand side)... Then one shot from there (pointing to 30 \leq and to the corresponding bar of the graph on the right) ... which ... that's why it's so small and 3 from there (20 \leq) ... and one from there (0 \leq) ... so, it's random basically.

39. Res: What will happen in the long run after many shots?

40. Nick: I think it will look like that (pointing to the modelling distribution).

Nick articulated that the throws were chosen randomly from the modelling distribution and emerged towards the modelling distribution.

Nick's comprehension of the random choice of the balls from the modelling distribution seemed more explicit because it was based on a gradual movement from the local views (individual bars of the data-centric and modelling distribution) to the global views of both perspectives of distribution, taking into account the shapes of the distributions as interdependent wholes. As a result, he very clearly articulated that in the long run the data-centric distribution would look like the modelling distribution.

Sarah managed little by little to come to an understanding of the random choice of throws from the modelling distribution; this comprehension appeared more and more explicitly in her words:

41. Sarah: After the highest amount that you put in. It depends really because every time he takes a shot ... it's just random, but it could be any one of the values he's chosen, but it might not be ... so, if it is then, it calculates it, and at the end you should get your values on the graph.

The above statement is indicative of Sarah's intuitive understanding of the existence of a calculating process that denotes the mathematical model. We (as expert observers) can relate her reference to the calculating process as connected to our understanding of the mathematical models (theoretical distribution) in which we attribute probabilities to a range of possible discrete or continuous outcomes.

Nick adopted the notion of the calculating mechanism that was introduced by Sarah and added:

42. Nick: After a certain amount of shots, the graphs would look the same because they are the same thing. This is calculating the end results (graph on the left) and this is doing it, while it is happening (the graph on the right) ... That's like the future and that is like the present.

43. Sarah: That's the overall.

44. Res: Which one is the future and which one is the present?

45. Nick: This is like the future (pointing to the graph on the left hand side) and this is what's gonna happen and this is what is happening (pointing to the graph on the right hand side)(Note 2).

Looking at the graphs, Nick deduced that “after a certain” number of throws being chosen from the modelling distribution, a data-centric distribution the same as the given modelling distribution would result. Indeed, he very clearly articulated that the modelling distribution displays the “future” outcome and the data-centric distribution displays the results of the present mechanism at play.

By referring to the end results it seems like he is using a target interpretation. So, in holding both an intention view and a target view, he is close to the coordinated view.

We see how the *BasketBall* simulation enabled them to deal at the same time with an empirical distribution (which I called the data-centric distribution) as a whole, while adopting a global viewpoint of their data and the mathematical model, which is the modelling distribution. At the very end of their session:

46. Sarah: So, that's calculating what each shot is gonna do (the graph on the left hand) until it reaches the point you told it to reach and then all the graphs would be the same.

Sarah appeared to be instinctively convinced that when a data-centric distribution would reach a certain number of throws, then it would be the same as the modelling distribution. As we witnessed in students' responses, this intuitive understanding of the “calculating mechanism” allowed some beginning to the bidirectional link of co-ordinating the two perspectives on distribution. Sarah and Nick appeared not only to have a strong sense of the modelling distribution as a target—suggesting a connection was being made from the data to the probability distribution—but also understood the role of large numbers in the resemblance of the two perspectives.

4. Summary

The process by which the students constructed connections for bridging the modelling and the data-centric perspectives on distribution is summarized in Table 1. The first column indicates the connections made by each student during the phase of manipulating the interface control points (ICPs; task 3). The second column refers to the second phase of manipulating the frequencies (task 4).

This paper is only presenting two cases, and while it does not present those cases completely, it does present those cases in detail.

Table 1. Categorisation of how the two pairs of students connected the modelling and the data-centric perspectives on distribution

Name of students	Manipulating the ICPs	Manipulating the frequencies
Sarah and Nick	T	T and I_{st}
Anna and James	I_G	T

When students manipulated the ICPs, Anna/James appeared to perceive the modelling distribution as the intended outcome and the data-centric distribution as the actual outcome. They had a sense of a general intention (I_G) when they talked about variation. Their articulations were explicitly characterised by the absence of a strong sense of the probabilistic mechanism and there was no progressive articulation of intuitive relationships leading gradually to a clear probabilistic-type language for talking about randomness and probability. Their reasoning about intentionality remains insufficiently clear, with at least two different possible interpretations: a) the intention is simply an expression of the pre-programmed deterministic nature of the computer - at least in their experience, or b) intentions are reflected in the actions of a modelling builder.

Stochastic intentionality (I_{st}) was articulated by Sarah and Nick (Note 3), who seemed to see the modelling distribution as the intended outcome, progressively generating the data-centric distribution. It can be concluded from their various references to chance that this perception of intention is probabilistic, and underpins the idea that the modelling distribution precedes—indeed generates—the data-centric distribution. Hence, they referred to the computer ‘wanting’ to throw the ball at various angles, a sentiment which we characterise as the situated abstraction: “the more the computer wants to throw at a particular angle, the higher is that angle's bar”.

Sarah and Nick made the reverse connection—from data to the modelling distribution—by perceiving the latter to be

the target to which the former is directed, without understanding the probabilistic mechanism by which the throws were selected randomly from the intervals of the modelling distribution.

When students manipulated the frequencies, we observed Anna/James and Nick/Sarah constructing a target interpretation (T) of the connection between the perspectives on distribution. During this phase, Nick and Sarah's comprehension of the random choice of the balls from the modelling distribution became more explicit. At the end of the session, their intuitive understanding of the probability distribution as a "calculating mechanism" led them to perceive the modelling distribution as the intended outcome (I_{st}) and the data distribution as the actual outcome, when an unknown finite number of throws would be randomly generated from the modelling distribution. Indeed, Nick very clearly articulated that the modelling distribution displays the "future" outcome and the data-centric distribution displays the results of the present mechanism at play.

The fact that these students articulated both types of connection suggests that for those who did not coordinate the two perspectives, future work on similar tasks might lead to a successful co-ordination in the same way Sarah and Nick successfully co-ordinated target and intention connections.

5. Discussion

When students constructed a general Intention interpretation (I_G), they appeared to recognize a movement that could be schematized as depicted in Figure 11:

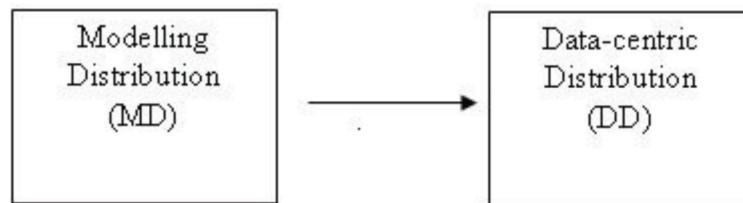


Figure 11. The general intentionality route

In the intentionality interpretation, students perceived the modelling distribution (MD) as the intended outcome and the data-centric distribution (DD) as the actual outcome, suggesting a connection being made, in which the modelling distribution in some sense generates the data.

General intention has the potential to become stochastic intention (I_{st} ; Figure 12) when randomness becomes a part of the interpretation of the student.

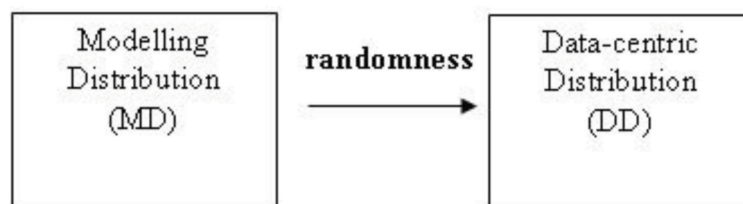


Figure 12. The stochastic intentionality route

Students could also interpret the modelling distribution (MD) as the target (T) to which the data distribution (DD) is directed (Figure 13).

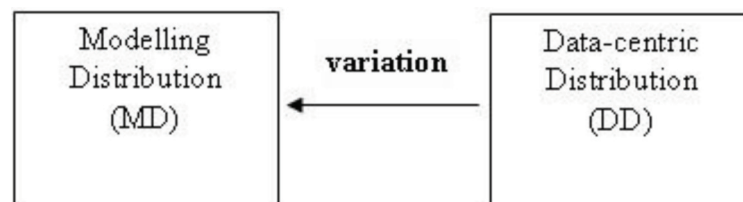


Figure 13. The target route

Schematisation of the stochastic intentionality and the target models lead us to a tentative model (Figure 14) for the connection of the data-centric and modelling perspectives on distribution

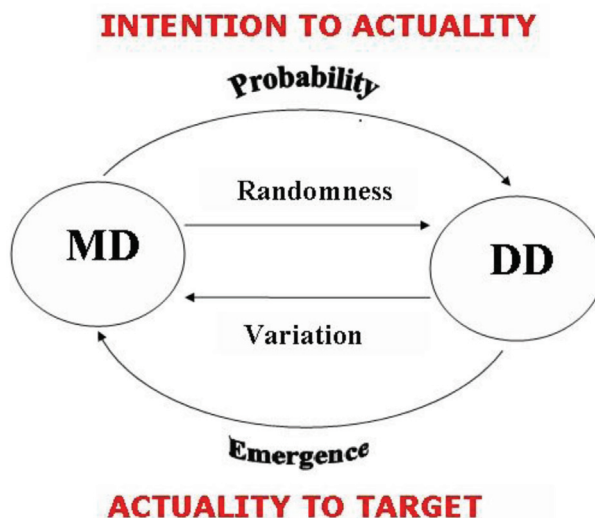


Figure 14. A tentative model for the connection of the data-centric and modelling perspectives on distribution

First, consider the connection from the modelling distribution (MD) to the data-centric distribution (DD), in the top half of Figure 14. Students typically tended to gravitate towards simple causal explanations when they observed the modelling distribution and perceived it as generating the data. As soon as they discovered the random mechanism that underlies the generation of the throws, their reasoning sought to assimilate this unexpected obstacle to employing causal explanations. In place of deterministic causality, they substituted the quasi-causal actions of agents to operationalise randomness.

Now consider the connection for DD to MD in the bottom half of Figure 14. Many students first recognise variation in data and make a target interpretation—a connection from the data-centric to the modelling distribution.

Students had great difficulty, however, in operationalising variation to explain this connection, and turned to relatively vague references to emergence, as if emergence were a causal agent. Quasi-causal emergence becomes, therefore, the driving mechanism by which the data-centric distribution targets the modelling distribution.

Some students made an intuitive synthesis of the *modelling* and *data-centric* distributions when they constructed an intentionality model that was dependent upon a strong appreciation of quasi-causal probability and a target model that was dependent upon the recognition of quasi-causal emergence.

The target and intention models are perhaps not dissimilar from how experts appreciate the co-ordination of the two perspectives on distribution. However, experts add to this image a co-ordinated understanding of how the probabilistic and emergent mechanisms are related through the Law of Large Numbers.

5.1 Connecting the Two Epistemological Perspectives on Distribution

Piaget and Inhelder (1951/1975) portray randomness as inconceivable within operational thinking. At the formal operational stage (beyond 11 years) of development they suggest that the understanding of randomness is resolved by the invention of probability.

In this study, it seems that the students (Note 4) used an idea of the quasi-causal actions of agents to operationalise randomness in a parallel manner to how Piaget and Inhelder claim that the human mind invents probability to operationalise randomness (1951/1975). In one direction—from the model to the data—is a sense of intention, which the students attributed variously to the human modeller or the tools within the software that activate the random generation of the data. Those tools within the software enable students to reason about the generation of the data-centric distribution in terms of cause and effect relationships of the actions of agents (ICPs) on behalf of the modelling distribution. The ICPs represent in a situated way the parameters of a modelling distribution and so potentially offer themselves as situated roots of probability theory. In that sense, probability is itself a causal-like agent that enables the modelling distribution to generate the data distribution.

It is possible to identify the distribution from the data and to conceive of a downward causation (Rockwell, 2007), or “macrocausation,” that means the modelling distribution has causal powers over the random selection of throws from which the data-centric distribution is constructed. However, according to Rockwell (2007), downward causation “can only happen if those macroscopic entities possesses (sic) what are called emergent causal properties” (p. 2). In our case, the macroscopic entity, the modelling distribution, does have emergent causal properties since, as we see with the target interpretation, the modelling distribution can be thought of as emerging from the data.

When the students in this study tried to make the connection from the data-centric to the modelling perspective, they observed the variation in the data, out of which the modelling distribution emerged. Acknowledging the obscurity of ‘cause and effect’, emergence itself is seen as a causal-like agent that gives rise to the modelling distribution as an emergent phenomenon arising out of the data-centric distribution.

Students cannot entirely predict what a data-centric distribution will evolve into because the modelling distribution is a complex system and, like all complex systems, it provides students with a mystery. Their limited understandings of probability and emergent phenomena as mathematical modelling constructs are quite evident. Students expressed unstructured interpretations that could be seen as rooted situationally in the abstracted theories of probability and emergence. They also exhibited profound difficulties in verbally describing and interpreting the quasi-causal agents of probability and emergence. In the target connection, students’ limited ability to express more than a superficial idea about emergence relates, we presume, to a lack of experience of the sorts of issues embedded in complexity theory.

Ultimately, some students make an intuitive synthesis of the modelling and data-centric perspectives when they construct an intentionality model that is dependent upon a strong appreciation of quasi-causal probability as well as a target model that is dependent upon a strong recognition of quasi-causal emergence. Issues of how the students create this intuitive synthesis are discussed in greater detail in Prodromou (2008).

5.2 Research Implications

The most significant issue that emerged from this study is the way of coordinating the two epistemological perspectives on distribution. As the researcher has discussed in the above sections, students come to a better understanding of distribution when they can construct both the intentionality and target models and construct connections between them.

In future research it will be fruitful to understand better how, when, and why such models are formed or acquired, and how they interact with the ongoing development of the ideas of randomness, quasi-causal probability, variation, and quasi-causal emergence.

In this research study, students had great difficulty in operationalising the target connection and turned to relatively vague references to emergence, as if emergence were a causal agent. There is a research need to support students to overcome verbal and conceptual limitations in articulating ideas about quasi-causal emergence and to help them understand, use, and communicate about these relevant abstract constructs.

Future research should aim also to develop new conceptual tools, both heuristics and qualitative tools, to help learners construct intentionality and target models for connecting the two perspectives on distribution. It is important to help students think about those models in new ways and subsequently to explore how learners can develop awareness and understanding of the relationship between the intentionality model and quasi-causal probability, the target model and quasi-causal emergence.

The researcher acknowledges, at the same time, the possibility that technological tools are playing a vital role in the necessary research into the development of understanding, and in providing fruitful learning environments. Learning environments like *BasketBall* simulation might support this intuitive synthesis of the two epistemological perspectives on distribution, with its dynamic features:

- Deterministic and stochastic types of causality.
- The sense that there is something playful about the environment.
- The environment presents the MD as that which generates the data and the DD as that which emerges into the model.
- Graphs are made available to examine and scrutinise both perspectives on distribution.

The central design feature of the *BasketBall* simulation was the way in which the researcher began with students’

known understanding of determinism, and empowered them to use this idea of causality to construct the intentionality interpretation, dependent upon an appreciation of quasi-causal probability. The target interpretation was unexpected but there is fascinating research to be done in discovering more about how learners interact with such software, and how their interactions affect the construction of the quasi-causal agents when other materially based methods are deployed. There is also a need to explore further the cognitive developmental intentionality and target models, to see how those models apply to different settings, to validate those models, and to investigate how they may be used to promote distributional reasoning, thinking, and literacy through carefully designed instruction.

5.3 Teaching Implications

The existing literature offers teachers a wealth of advice about teaching stochastic concepts (both misconceptions and cognitive developmental processes of stochastic concepts), at the risk, perhaps, of overwhelming them. Ultimately, the contribution of this research will be to help teachers understand and provide students with a broader repertoire of causal models for understanding statistical concepts. It is vital for teachers to spend substantial time in their statistical courses in engaging their students in the nature of causality and how it behaves differently in stochastic settings. It makes sense that explicit conversation of such issues might be helpful to both students and teachers.

If such a focus on causality could be placed in curricula materials and if causality does still arise as a coordinating agent of the two perspectives on distribution, we may be justified in describing a bridging based on causality as a new way of learning and knowing, or of reconciling two traditionally different worlds: the deterministic and the stochastic. The researcher suggests that this is a promising avenue towards inducing a cognitive developmental construction of statistical concepts.

References

- Abrahamson, D. (2008). Bridging theory: Activities designed to support the grounding of outcome-based combinatorial analysis in event-based intuitive judgment - A case study. In M. Borovcnik & D. Pratt (Eds. of Topic Study Group 13, Research and Development in the Teaching and Learning of Probability), *Proceedings of the International Congress on Mathematical Education (ICME 11)*. Monterrey, Mexico: ICME.
- Abrahamson, D., & Wilensky, U. (2007). Learning axes and bridging tools. *International Journal of Computers for Mathematics Learning*, 12(1), 23-55. <http://dx.doi.org/10.1007/s10758-007-9110-6>
- Caput, M., Girard, J.-C., & Henry, M. (2011). Frequentist approach: Modelling and simulation in statistics and probability teaching. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-Challenges for teaching and teacher education: A joint ICMI/IASE study: The 18th ICMI study*, pp. 85-95. New York, NY: Springer.
- Cohen, S., & Chechile, R. A. (1997). Probability distributions. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education*, pp. 253-262. Amsterdam, The Netherlands: IOS Press and the International Statistics Institute.
- DfES. (2000). *Mathematics: The national curriculum for England*. London, England: Department for Education and Employment.
- Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.
- Konold, C., & Pollatsek, A. (2004). Conceptualizing an average as a stable feature of a noisy process. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking*, pp. 169-199. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- 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, 1*, pp. 176-194. University of Lisbon, Portugal.
- Noss, R., & Hoyles, C. (1996). *Windows on mathematical meanings: Learning cultures and computers*. London, England: Kluwer Academic Publishers. <http://dx.doi.org/10.1007/978-94-009-1696-8>
- Papert, S. (1996). An exploration in the space of mathematics educations. *International Journal of Computers For Mathematics Learning*, 1(1), 95-123. <http://dx.doi.org/10.1007/BF00191473>
- 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, England: Routledge & Kegan Paul (Original work published 1951).
- Prodromou, T. (2007). Making connections between the two perspectives on distribution. In D. Pitta-Pantazi & G. Philippou (Eds.), *Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education*, pp. 801-810. Larnaca, Cyprus.
- Prodromou, T. (2008). *Connecting thinking about distribution* (Unpublished doctoral dissertation). University of Warwick, England.
- 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.
- Robson, C. (1993). *Real world research*. Oxford, England: Blackwell.
- Rockwell, T. (2007, July 8). *A defense of emergent downward causation*. Retrieved from <https://users.california.com/msmf/causewb.html>
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Notes

Note 1. See <http://ccl.northwestern.edu/netlogo/>

Note 2. This is a succinct expression of the target interpretation.

Note 3. Other students articulated I_{st} (Prodromou, 2008).

Note 4. The researcher does not suggest that the students in this study are at Piaget's operational thinking stage because the order of the three operational periods is invariant and the age at which any given period starts and ends varies significantly. It is dependent, inter alia, upon the individual's intellect, experiences, and cultural background, and not all individuals will reach the formal operational stage.