LOCAL AND GLOBAL THINKING IN STATISTICAL INFERENCE

DAVE PRATT
Institute of Education, University of London
d.pratt@ioe.ac.uk

PETER JOHNSTON-WILDER
Institute of Education, University of Warwick
p.j.johnston-wilder@warwick.ac.uk

JANET AINLEY
School of Education, University of Leicester
janet.ainley@le.ac.uk

JOHN MASON
Centre for Mathematics Education, Open University
j.h.mason@open.ac.uk

ABSTRACT

In this reflective paper, we explore students' local and global thinking about informal statistical inference through our observations of 10- to 11-year-olds, challenged to infer the unknown configuration of a virtual die, but able to use the die to generate as much data as they felt necessary. We report how they tended to focus on local changes in the frequency or relative frequency as the sample size grew larger. They generally failed to recognise that larger samples provided stability in the aggregated proportions, not apparent when the data were viewed from a local perspective. We draw on Mason's theory of the Structure of Attention to illuminate our observations, and attempt to reconcile differing notions of local and global thinking.

Keywords: *Statistics education research, Task design, Informal statistical inference, Sample size, Local and global meanings or perspectives, Structure of Attention*

1. WHAT IS INFERENCE AND IN WHAT SENSE IS IT INFORMAL?

Statistical inference is typically introduced as a formal topic in the curriculum at around age 16 or older. Inferential analysis is typically taught as a tool for judging the source of variation in data. Students' lack of comprehension has been widely reported and in response there has been a recent research effort to understand how better to approach the topic from a pedagogic perspective. One response has been Exploratory Data Analysis, in which students attempt to infer informally about underlying trends in data without explicit reference to probability. These approaches make even more urgent a deeper understanding of how young students might make sense of this inferential process.

Statistics Education Research Journal, 7(2), 107-129, http://www.stat.auckland.ac.nz/serj

In this paper, we reflect on recent small-scale experiments to clarify understanding of young students' activity when engaged in an inference related task designed to trigger intuitions, assumptions and conceptual thinking. Our aim is to elaborate the conceptual struggle that needs to take place for young students to engage in inferential reasoning. In so doing, we acknowledge a constructivist stance in which we search for naïve conceptions (as opposed to misconceptions, a distinction delineated by Smith, diSessa, & Rochelle, 1993) that might serve as resources for further development. We begin by clarifying our perspective on what we see as the conceptual roots of statistical inference because it is in those roots that we may find informal inference.

Makar and Rubin (2007) give a working definition that we find useful: "We consider informal inferential reasoning of statistics in broad terms to be the process of making probabilistic generalizations from (evidenced with) data that extend beyond the data collected." For us, this definition describes, without misrepresentation, statistical inference per se. By focussing on the conceptual basis for inference, we can begin to imagine younger students engaging in such activity without necessarily conducting formal statistical hypothesis tests or building carefully defined confidence intervals, as we would see in a classical statistics course. Our focus is primarily on students around the age of 10 or 11 years, well before any formal teaching of inference. It is therefore likely that intuitions will not yet have been formalised through schooling and that student thinking can be thought of, in that sense, as informal.

However, we wish to probe a little further into what Makar and Rubin might mean by "making probabilistic generalisations from data." Immediately, we recognise a direction to the statement – from data to probabilistic generalisation. Inference is concerned with identifying patterns in the form of trends or statistical parameters in the "underlying" population. There has been recent research interest in how children might make, or be encouraged to make, inferences such as these by studying how children think informally about populations, given samples of data (Ben Zvi, 2006; Pfannkuch, 2006). Typically, the activity has been focussed on using sampled data to make statements about a finite population from which the data were drawn. However, other possibilities for informal inference exist. Our interest focuses on situations where the population cannot be described in terms of the total finite dataset but instead is most adequately described through a probability distribution. The data may be used to draw inferences about aspects of an infinite population or process. Indeed, our focus in this paper will largely be on inferences about the probability distribution associated with outcomes from a die. Here again it is important to distinguish between an expert sophisticated perspective on what was happening and the likely relatively naïve perspective of a 10-year-old. Although we might see the activity as being rooted in making inferences about a probability distribution, it is reasonable to think that 10-year-olds saw the game they were playing as trying to guess what the die looked like.

We also note that in some reported studies on informal inference the activity used may be open to ambiguous interpretation. During SRTL-4 in Auckland, New Zealand, the theme for discussion was students' thinking about distribution, rather than inference. However, several presentations showed students working informally with data. There was a realisation during discussion that the focus of the students may at times have been on the dataset as if it were the whole population, whereas the teacher's attention may have been on the underlying population from which that dataset was only a sample. We referred to these two situations as Game 1, where the dataset was all there was, and Game 2, where the dataset was merely a sample. This distinction is critical in considering how students think about inference. Game 1 allows no room for inference (*what you see is what you get*) whereas Game 2 demands an ability to make inferences about the population on the basis of the data available (*what you see* might be *what you get*). We designed a task involving making inferences about the probability distribution of a die, in which, even from the students' perspective of simply guessing the configuration of the die, the focus is firmly in Game 2. The results of throwing the die are a sample arising from a "hidden" (or "underlying" in the usual statistical parlance) generator (or "population").

Let us now return again to Makar and Rubin's definition, and consider the thinking involved in 'making probabilistic generalizations from data'. Prodromou (2007), and Prodromou and Pratt (2006) have shown that 15-year-old students, using a Basketball microworld, were able to make connections between data and what was called the modelling distribution. (We can think of the modelling distribution as a probability distribution used to model any particular phenomenon. Prodromou distinguished two perspectives on distribution, the modelling perspective and the data-centric perspective, which attends to the sample of data.) She has proposed the mechanisms that act in the explanations expressed by learners as quasi-causal agents connecting the modelling and the data-centric perspectives. These agents are not causal in the sense of direct cause and effect. Rather they are invented substitutes, which for learners play the role of causal agents given the lack of any explicit determining agent.

In the 'inferential direction', from a data-centric to a modelling perspective, some of Prodromou's students saw the modelling distribution as a target, towards which the data-centric distribution was aiming. There is a sense of variation in the data, out of which the population is an emergent phenomenon, like a trend (Prodromou, 2008). In the absence of any explicit cause for how this happens, emergence itself is seen as a somewhat mysterious causal-like agent that enables the modelling distribution to emerge out of the data-centric distribution.

In the opposite direction, from a modelling to a data-centric perspective, there is a sense of intention. In Prodromou's research, the intention is attributed by the students variously to the human modeller, the characters in the software or the tools within the software that trigger the random generation of data from the modelling distribution. Software tools allowed students to explain the generation of the data-centric distribution as 'caused' by the actions of agents on behalf of the modelling distribution. We would recognise such explanations as the situated roots of a view of probability theory as a causal-like agent that enables the modelling distribution to generate the data-centric distribution. Indeed, in classical statistics, data are seen as generated through a mixture of signal and noise from a modelling distribution.

Although we would regard making such connections between data and population as lying at the heart of informal statistical inference, neither probability nor emergent phenomena are trivial areas of mathematical modelling, and so, when we observe children, we are likely to see either (i) naïve understandings, which serve to make sense of the world in the absence of such theory, or (ii) meanings, which appear to us to be rooted in a situational way to these abstracted theories. Maker and Rubin (2007) refer to the importance in informal inference of aggregate thinking, sample size, controlling for bias, and tendency. Our focus is firmly related to aggregate thinking and how students might perceive the effect of sample size on the inferential process. In particular, we have become interested in the shifting of attention between what is happening in the here-and-now, the immediate, and what is happening in an aggregated sense over the longer-term.

2. THE LOCAL AND THE GLOBAL

Our interest in informal inference emerges out of aspects of previous work. Pratt and Noss (2002) reported that 10- to 11-year-olds were able to articulate expert-like views about short-term randomness, through meanings that were immediately accessible. These meanings were described in that study as *local*, in the sense that such resources focus on trial by trial variation. According to Pratt and Noss, local meanings refer to unpredictability, irregularity and uncontrollability.

In contrast, these students did not easily express meanings for long-term randomness. Using a design research approach (Cobb et al., 2003), Pratt built a computer-based domain of stochastic abstraction called *ChanceMaker*. This microworld provided 'gadgets', simulations of everyday random generators such as coins, spinners and dice, whose behaviour was controlled through a 'workings box', which is an unconventional representation of the distribution. Figure 1 shows the default workings box for the dice gadget. The software allowed children to explore the gadgets using tools to display results in different ways (pie chart, pictograms, and lists), and to examine and edit the contents of the workings box, which controlled the behaviour of the gadget (see Figure 1). Challenged to identify which gadgets might not be working properly, the children began to use the tools provided to decide how to mend the gadgets.

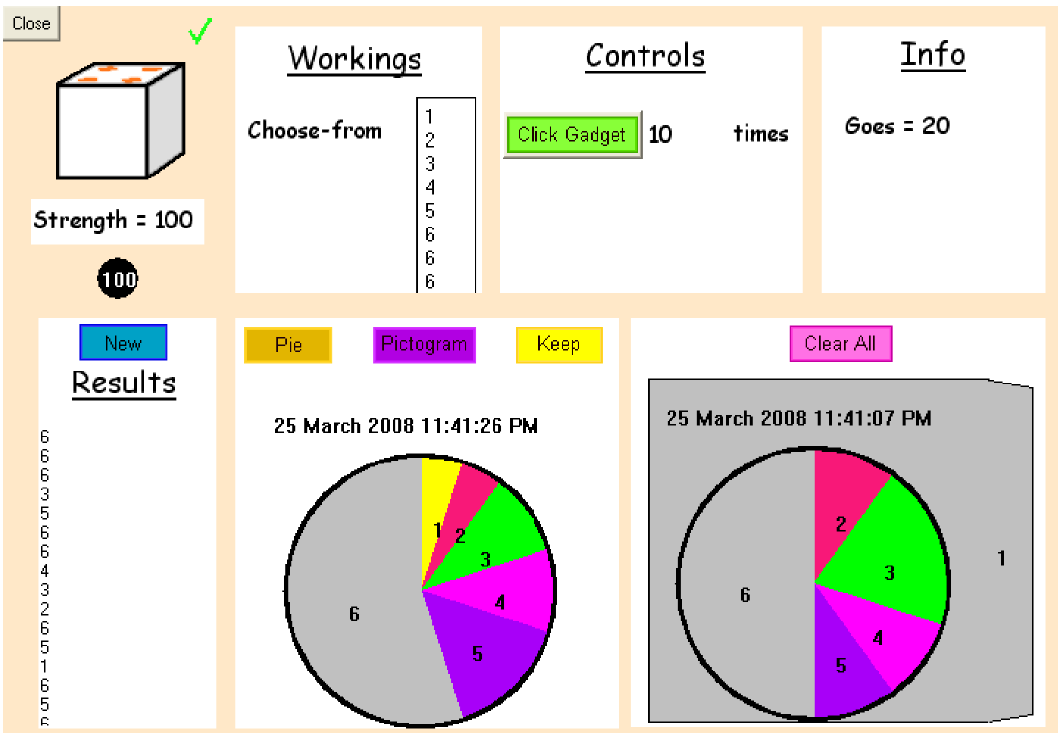

Figure 1. The ChanceMaker dice gadget can be opened up to reveal a "broken" workings box. Here the student has created a pie chart from 10 throws, kept that picture, and then created a second pie chart from 20 throws.

Gradually, the children began to articulate meanings for long term randomness, which focussed on an aggregated overall view of the stochastic, such as *the proportion of outcomes was predictable* (probability), *the proportion of results stabilised when more* *trials were executed* (large numbers), and *the observer is able to exert control over these proportions through manipulation of the possibility space* (distribution).

Pratt (2000) noted an interesting correspondence between local and global meanings:

> "…local resources tend to be inverted in relation to their global counterparts. Thus, unpredictability as a local resource is inverted in comparison to the global resource of predictability (in a proportional sense). Similarly, control cannot be exerted locally whereas there is a global resource for control through manipulation of the distribution." (p. 609)

In Pratt's study, the students were observing patterns in the data via the different representations. By interacting with the workings box, they were in effect creating a probability distribution to generate data. Their task became one of understanding the nature of the control exerted by the workings box over the variability and the structure of the data. It is clear from this work that it was far from trivial for students of this age to identify the global structures that enable statisticians to view aggregated results as predictable amidst the local variability of the data. Nevertheless, the design of *ChanceMaker* appeared to support students in articulating situated heuristics such as "the more times you throw the dice, the more even is the pie chart," and to recognise that such a heuristic could be applied to explain the behaviour of various gadgets.

Whereas Pratt's approach led to students initially articulating local meanings before the emergence of global meanings alongside the local, Johnston-Wilder (2006) has shown how the perceptions of students of various ages (the youngest were three years older than those in Pratt's study) shifted back and forth between the local and global, as they attempted to identify whether various types of dice were fair or not. In one-to-one interviews, students rolled each die and recorded the outcome, pausing frequently to reflect on the observed sequence. As the outcomes unfolded, students' attention shifted between a local perspective, in which successive outcomes were seen as disordered and unpredictable, and a global perspective, in which they looked for an empirical distribution of the outcomes. This shift of attention was sometimes rapid and often subtle. In the light of Pratt's study, we noted the direction of the shift with particular interest.

In Johnston-Wilder's study, students were trying to make judgements about the fairness of various different kinds of dice by looking at the observed outcomes. A student's attention was initially focussed in the local perspective, typically looking at short sequences of outcomes, and often seeking patterns in these. Although such patterns might appear by chance, they were typically not sustained as more outcomes were generated. Students often discerned sequential patterns, which they thought might be extended in later outcomes, using them to make predictions about future outcomes. However, these illusory patterns appeared to be a significant distraction for the student trying to make a judgement about whether the dice might be considered to be fair.

Looking at the outcomes through the global perspective involves looking across the space of possible outcomes to consider the frequencies of the various outcomes. The focus of attention is not on successive outcomes, but on the distribution of the outcomes across the outcome space. To see the process through the global perspective involves a different way of looking at the outcomes, and through this perspective it is possible to see an emergent order and pattern in the distribution.

However, the distribution of outcomes can be thought of in two different ways. Firstly, the distribution can be a theoretical probability model which expresses what one expects from a process. Secondly, the observed distribution of outcomes might be viewed empirically as a representation of the underlying probability model. In the case of a student experimenting with, say, a spherical die (a hollow sphere with a hidden weight moving around inside the shaped interior such that it comes to rest in one of six different orientations), the student might expect the die to be fair, beause the spherical appearance gives the die an apparent symmetry. The student might therefore hold in mind a theoretical distribution according to which they expect each outcome to occur equally often. Working with this theoretical prior model of a distribution, such a student might try consciously to control the process of throwing the die to produce an outcome that had not yet been observed in the outcomes so far. In doing so, the student has shifted from the global perspective, in which they were comparing the observed frequency distribution of outcomes with their mental model of what they expected the distribution to look like, to a local perspective in which they look for a particular outcome to occur next.

In interviews with students as they experimented with the different dice, Johnston-Wilder observed that the students' attention shifted between, on the one hand, the unpredictability of the next outcome and the lack of order and pattern in short sequences of outcomes, and on the other hand, the order and pattern that was seen to emerge in an empirical distribution. When the student tried to infer the distribution from small samples, the apparently conflicting information arising from successive small samples appeared to lead the student to make particularly rapid shifts of attention between global and local perspectives.

There are two interesting contrasts between the use of global and local by Pratt and by Johnston-Wilder. Whereas Pratt's work focussed on the emergence of the global out of the local, Johnston-Wilder's observation was one of constant shifting of attention between the local and the global. Secondly, in Pratt's work, the students had available to them the workings box, which came to be seen as a representation (of distribution from our perspective) that could be used to predict behaviour and results, whereas Johnston-Wilder's students could see the dice but did not have access to the associated probabilities. Pratt's students could therefore draw on information about the workings box as well as the generated data, whereas Johnston-Wilder's students had available the data and whatever prior distribution they held for the dice, in the sense of expectations about how it might behave.

This difference could be important because inference is more closely related to the activity of Johnston-Wilder's students. When statisticians make inferences, they attempt to make descriptions of the population (the distribution, or at least statistical parameters such as the mean, which describe elements of that distribution) based on data that have been sampled. Johnston-Wilder's students, in trying to understand what was happening when the die was thrown, were trying to infer something about the underlying infinite probability distribution (at least from our perspective), in some respects the inverse of what Pratt's students were doing. By attending to the workings box, a representation of the modelling distribution, and observing consequential changes in the data, Pratt's students might be expected to see, in Prodromou's (2007) terminology, intention in how the human or computer agent generates the data through the modelling distribution. In contrast, again using Prodromou's terminology, one might expect Johnston-Wilder's students to see the data as targeting the modelling distribution in the way that the model emerges out of the data. However, it is unclear whether the shifting of attention observed by Johnston-Wilder could simply be accounted for by differences in theoretical perspective from that of Pratt, or by methodological issues related to the task differences, whereby target connections are more prone to such shifting than intention connections.

In this paper, we seek to elaborate on the shifting between the local and the global in such a way that we begin to recognise the difficulties that students may have in making informal inferences that connect data to probability distribution. Our approach will be to examine fresh data, arising from a small-scale study in which students were challenged to infer the nature of a single hidden *ChanceMaker* die, given data being generated by the students' manipulation of that die. We aim, through this new elaboration, to guide the design of new resources that would have the potential to support young students' inference-making. In this respect, this paper could be thought of as an extended reflection, marshalling thoughts geared towards the next phase of a long-term design experiment. We have found it useful to extend our corpus of data from the work of Pratt and Johnston-Wilder with some fresh data, collected in an attempt to tease out the subtle differences alluded to above. In the next section, we explain further the basis of the additional data which will inform our understanding of the relationship between the local and the global in statistical inference.

In our discussion of the new data in this paper, we have drawn upon a framework for the Structure of Attention (see, for example, Mason & Johnston-Wilder, 2004). This describes how a person's attention can shift rapidly between different foci, related to different ways of attending. Mason and Johnston-Wilder refer to the following five ways of giving attention to a situation:

- attention on the whole, the global;
- attention on distinctions, distinguishing and discerning aspects, detailed features and attributes;
- attention on relationships between parts or between part and whole, among aspects, features and attributes discerned;
- attention on relationships as properties that objects like the one being considered can have, leading to generalisation;
- attention on properties as abstracted from, formalised and stated independently of any particular objects, forming axioms from which deductions can be made.

(Mason & Johnston-Wilder, 2004, p. 60)

There is considerable evidence that people working to make sense of random phenomena are sometimes attending to what will happen next (immediately), to what has just happened, and to what has been happening over a longer period. They may easily circle around these very rapidly, or focus for a time on one or the other. It is a reasonable conjecture that some people at least are *seeking a relationship* between two or three of these, not always with success. But the overall goal of the instruction is that they see these relationships as examples of properties that can hold in other situations.

Children in the data discussed below run into conflict with the essential unpredictability in the short-term, and the fact that only long-term statistics (summaries of data) are likely to show some relative invariance. Sometimes the children are *discerning details* in the graphs for example (past history) and *seeking relationships* between the past history and the number of occurrences of something. But, in the back of their minds there is a nagging doubt because these 'relationships' are at best approximate.

One aim of the tasks is to get learners to restructure their attention from the local to the global, from the details of specific events to the summary statistics of a large number of events. This could be described as *perceiving a property* which is instantiated in the data collected. However, this has an extra wrinkle because summary statistics are never exact, only approximated by instantiations.

In the discussion of the data that follows, we use this way of thinking to *account for* the way that the children often show a somewhat tenuous desire for more trials. Sometimes it is the interviewer who suggests the need for more trials, and sometimes the idea comes from the children. We suggest that, in this study, children seem to be seeking relationships between three distinct features of the software: the 'workings box' (whether they have access to it or imagine it); the role of a 'large number of trials'; and the graphical representation of the outcomes (whether it be 'pie-chart or pictogram').

3. A FURTHER APPROACH TO STUDYING THE LOCAL AND THE GLOBAL IN INFORMAL INFERENCE

Building on the work of Prodromou and of Johnston-Wilder, we designed a small-scale study, which aimed to explore children's thinking-in-change (Noss & Hoyles, 1996). One aspect of the study was to examine in detail students' attention to the local and global when they worked with *ChanceMaker* as in Pratt's original study. However, in order to connect this work to the second aspect of the study, and because of our limited resources, we restricted the children to work only with the die gadget.

For the second aspect, we wished to explore a situation which seemed to bear the hallmark of Johnston-Wilder's work, in that the students were not aware of the underlying probability distribution, and yet had the advantages of simulation, such as the ease of generating large amounts of data as available in the *ChanceMaker* study. In fact, we recognize that in pedagogic tasks which aim to engage pupils in inference, there is a potential gap between the focus of the teacher/designer on reasoning about the population (looking at the population through the sample as in Game 2 in Section 1) and the focus of pupils who may be looking at and reasoning about the sample, without understanding the true nature of the game (as in Game 1 in Section 1). We therefore wanted to design a context in which it would be clear to the pupils that they needed to attend to the sample *in order to* reason about the population. We therefore implemented a simple modification to the original software, which allowed the 'workings box' for the die to be hidden (see Figure 2). Using this modified software (which, for the sake of clarity, we call *InferenceMaker*), we designed a task based on the creation of a 'funny' die which children could explore in order to guess what numbers were on its sides. The software allowed the user to enter any numbers into the workings box, and also any number of numbers, so it was possible to 'make' very unusual dice.

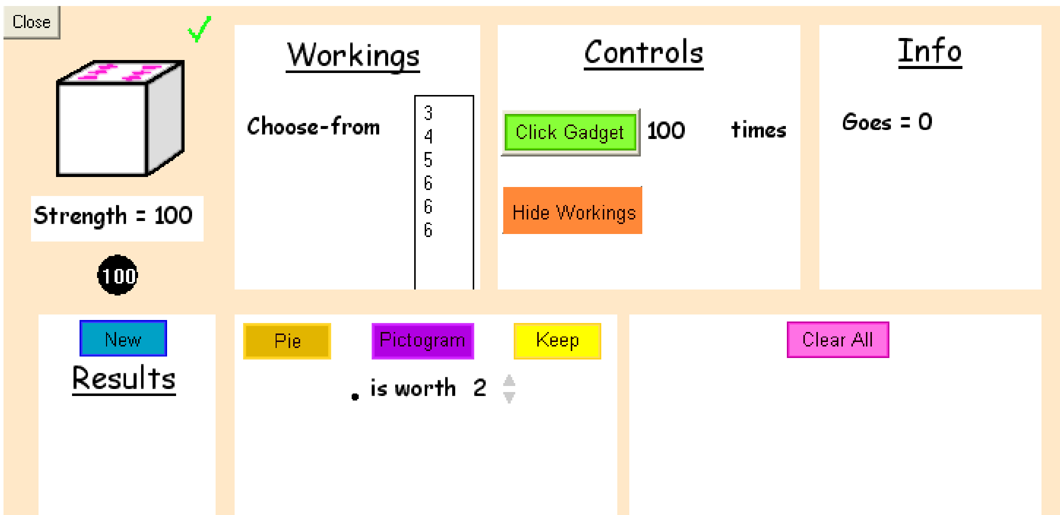

Figure 2. In InferenceMaker, it was possible to edit the workings box and then hide it by clicking the "Hide Workings" button. Children were then challenged to infer the configuration of the die by generating data and charts.

We conducted clinical interviews with small groups of 10- to 11-year-old children, in the final year of primary school. We worked with children from a single class group, covering a range of attainment. While the children were working, Camtasia™ software was used to capture all activity on the computer and audio recordings were made of all discussion. Field notes were kept by the researchers.

When working with *InferenceMaker*, after a brief introduction to the software, a funny die was created in secret by editing the workings box. This was then hidden, and the children were challenged to guess what the new die looked like, using any facilities of the software that they chose. To focus their explorations, they were told how many sides the die had (i.e., how many items were in the workings box). Once they had decided on a description, a discussion was held about how the description was arrived at, before the workings box was revealed.

In the ensuing elaboration, we shall first discuss students' attention on the local and global when the workings box was available, as in the original *ChanceMaker* study, and secondly two main themes relating to how the students' attention to the local impacted upon their inferential reasoning when the workings box was hidden. Our focus on attention throughout Sections 4 and 5 is informed by the Structure of Attention framework.

4. THE LOCAL AND THE GLOBAL WHEN CONNECTING THE WORKINGS BOX TO THE DATA

We report first on the activity of Jim and Ivan, as they worked on mending the die gadget (as in the *ChanceMaker* study) for 30 minutes. The boys were presented with a 'broken' die gadget, whose workings box contained too many sixes (see Figure 1). The workings box was visible, but at first the boys paid little attention to it. Most of the initial interaction with this gadget was done by Jim. He approached the investigation systematically, ensuring that each batch of data had exactly the same number of trials. During this work, Ivan was very quiet and said almost nothing. After several experiments with 32 trials in each, Jim reported that "the dice might be a bit wonky if it keeps having six as the most, like the biggest." However, although he had stated clearly that the die was 'wrong', he later commented that "it could just be total luck that it happens to do that.".We continued to press the two boys to suggest ways that they could be more certain about whether there was something wrong with the die gadget, but the boys' only suggestion was to experiment with the 'strength' control, to see whether this made a difference. At this point, Ivan began to contribute more, and after a further set of 32 trials, he quickly (and correctly) stated, "It doesn't make a difference." Jim agreed that this suggested that "there might be something wrong with the dice."

The persistence of the relatively high frequency of sixes over several sets of trials attracted Jim and Ivan's attention. They attended to the detail of the proportion of sixes in each set of trials, and related the frequency of sixes to the frequencies of the other outcomes. The details became, for the boys, a property, which Jim has expressed as "something wrong with the dice." The boys compared what they noticed with their unarticulated expectation that the distribution of outcomes should be uniform.

The boys then turned their attention to the workings box, and Ivan immediately identified that there were "loads of sixes in there," confirming their conjecture. He edited the contents of the workings box, deleting two of the three sixes, to leave one of each digit from 1 to 6. When the boys went on to test the mended die, Jim again took control of the experiments, but the uneven outcomes from the 32 trials left him unconvinced that the die was truly mended.

For clarity in all protocols we have adopted a convention in which we use the word, such as 'six', to represent the cardinal number on the face of the die or in the workings box, and the digit, for example ‘2’, to represent the frequency with which an outcome appears in the workings box or in the results.

Res: So what do you think now? Do you think the dice is mended or do you think it’s still wonky?
Jim: Well, dices aren’t supposed to do equal amounts of each number, really. Cos they’re dices, but… it’s… it doesn’t give out that much, but… it’s not like having six as the biggest thing any more… We could probably fix it properly if we changed the numbers round on the workings a bit more. And we could make it equal I suppose.
Res: How would you do that?
Jim: Umm… If there were less ones, you could put 2 ones in. Or something. But that would increase its chances by two and then that would make it go wonky, so…

In contrast, Ivan saw no difficulty with the conclusion that the die was now mended. He was certain that the die was working properly, “Because the workings there, everything’s good.”

Res: Oh, right! But it’s not coming… the numbers aren’t all coming out the same amount. Does that matter?
Ivan: No. Because in real life each one will sometimes be higher.

Even after they had altered the number of sixes in the workings box, the proportion of sixes observed in the subsequent set of 32 trials was still subject to the vagaries of random variation; 32 trials is still a relatively small sample upon which to base a judgement. There is no evidence here that the boys have yet perceived the wider and more powerful property relating to large samples: the Law of Large Numbers. Jim in particular appeared to expect more stability than he found from his relatively small samples of 32 trials.

5. THE LOCAL AND THE GLOBAL WHEN INFERRING FROM THE DATA

In this section, we consider how students worked with *InferenceMaker*, where the workings box was hidden. The children engaged enthusiastically with the task. They typically began by making single throws of the die and looking at the results list, but, with a little prompting, they moved to using the pie chart or pictogram to make predictions about the die as they extended their sample. Their final samples varied in size considerably, and there was no general recognition that a larger sample might be more effective. We shall say more about this in the next subsection, but first we want to discuss the dominant tendency for students to attend to the local.

A recurring feature of the children’s activity was their focus on the changes which occurred in the appearance of the graph as they grew their samples by adding more throws, and the relative invisibility of more stable features.

Rob and Carl had been given a ‘funny’ die with six sides [3 4 5 6 6 6]. They decided quite quickly that the only numbers on the die were three, four, five and six. They also realised that six was occurring most often, and so conjectured that there were 2 sixes (‘the six is 2’). This left them with four faces on the die to fill. As their sample grew, three and five seemed to appear more often than four (though not as often as six). As they added more throws, and studied the pie charts, they explored different ways to describe the die. This conversation took place when they were examining the pie chart, generated from 130 throws (see Figure 3).

Rob: The six is 2 [sides].
Carl: Six is definitely 2 or 3. Three is 1, I think. No four is definitely 1. I think three is…
Rob: It's either three or five, isn't it [that is 2 sides]?
Carl: Three or five, may be 2.
Res: Why do you think six is 2?
Carl: Because it's getting the most rolls.
Res: Ah, why do you think it's 2 and not 3?
Carl: We don't, I think it's either 2 or 3.

130 throws

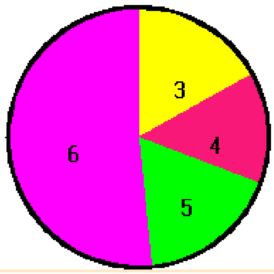

Figure 3. Rob and Carl make informal inferences after 130 throws.

At first sight, it seemed puzzling that the boys did not 'see' that six was taking up a much larger slice of the pie than either three or five, and so could not have been representing the same number of sides of the die. However, their extended discussion showed how their attention was, at this point, focussed on the way each slice was changing as the sample grew. The five-slice seemed to be getting bigger, but was it bigger than the three-slice? Their attention was focussed upon the way that each slice was changing and on the need to fill the appropriate number of 'faces' on the die: They recognised that if there were 2 sixes, and only 4 numbers in total, there must be another duplicate number somewhere on the die.

Throughout their work on the task, Rob and Carl seemed willing to grow their sample with larger numbers of throws, but whether in the hope of seeing stability in the results or simply because this was the only action available to them other than to make a conjecture and settle upon it, we cannot tell. The local and sense-triggered phenomenon of change seems to be at the forefront of their attention, whereas invariance (stability) is hard to detect as a relationship, particularly so because each new graph replaced the previous one as more throws were added. They gave no evidence (as we might have wished) of *perceiving a property* (global) of 'settling down in the long run'. In the end they had a sample of 280 throws before agreeing on their final (correct) description of the die.

We then challenged their confidence in this description by making a new sample of 10 throws. Their confidence proved to be fragile (Figure 4).

Carl: Six is the most
Res: What do you think?
Rob: 2, 2, 2 (pointing to the three sections)
Carl: No
Res: It's the same dice
Carl: 2 sixes, 2 threes and the rest are 1
Rob: Yeah but remember there's a four as well 2 sixes, 2 threes
Carl: 2 sixes, 2 threes and then 1 is five and 1 is...
Res: But you were saying to me that there were 3 sixes and 1 three, 1 four and 1 five
Carl: This is confusing!

10 throws

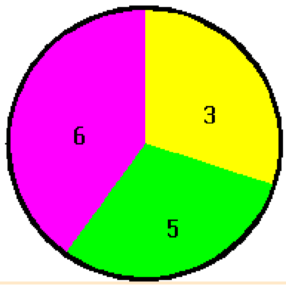

Figure 4. Carl and Rob respond to our challenge based on 10 throws.

Adding ten more throws to the sample produced yet another image, with a small slice for four and a reduced slice for five. When asked which image they believed. Carl favoured the one for twenty throws: "Because the 280 was just getting too stupid, I think, and had too much in.” Rob set about making a sample of 140 throws, half way to 280. Carl’s dismissal of the larger sample as “too stupid” seemed at odds with his earlier willingness to create more data, but may have reflected his frustration that more data seemed to produce more change rather than more stability. Carl’s dismissal of the larger sample might be explained in terms of the way senses work; the senses may have been activated by change and the learner’s attention may have been drawn to change. Although Rob and Carl were expecting invariance in the midst of change, as they may have learned to expect from their past experience of mathematics and from learning to make sense of their experiences in daily life, their attention was drawn to change rather than to invariance. They lost sight of the significance of the global perspective (the results of the larger sample of 280 outcomes) as the change arising from what had happened most recently in the smaller sample of 10 outcomes (a local perspective) had attracted their attention.

It was unclear what Rob and Carl were expecting or hoping to see as they added more and more throws to their sample. We conjecture that they were expecting to reach a point where the new, larger sample produced a graph which looked the same as the previous one; that is, where there was no change. However, as even small changes were grabbing their attention, because this is what our senses are most attuned to, they became frustrated. Even though they had drawn an initial inference from a sample of 280 outcomes, their attention was drawn by the change apparent in the subsequent sample of 10 outcomes. They did not have a secure perception of the (global) property that large samples were more informative than small samples, and perhaps assumed that the strategy of taking larger samples was the wrong one.

We now return to the work of Jim and Ivan, whose first encounters with the workings box in the software were described and discussed in Section 4. In what follows, we had used *InferenceMaker* to set up a 12-sided die whose workings box was hidden and contained the following: [1 2 2 2 3 3 3 4 5 6 6 7]. Jim and Ivan were told only that the die had 12 faces. Jim initially took the lead in controlling the software and generating data until they had a pictogram of 100 outcomes (Figure 5). Ivan commented from this graph that there were “loads of twos and threes,” but Jim found difficulty in interpreting what this graph implied about the die.

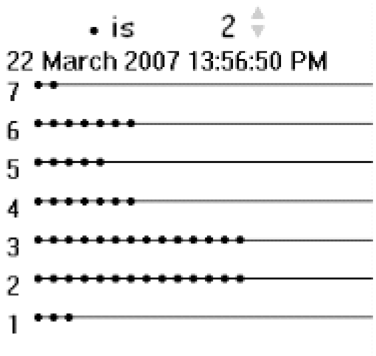

Figure 5. Jim and Ivan generated a pictogram based on 100 throws.

Ivan: Well last time it was all different (*indicating the outcomes*), and it was the same numbers (*indicating the workings box?)*

Res: So what might it have?

Jim: Umm. It might have 2 sevens and 3 ones and … (pause) and 7 sixes…

Jim seemed to have switched to describing the numbers of spots in the pictogram as though they were exactly the numbers of each element in the workings box. This appears to be an unintended metonymy, in which Jim's attention has switched from one detail to another. In contrast Ivan appeared to have a clear idea of how to interpret the pictogram.

Res:	Which number do you think that there might be most of on [the] dice?
Ivan:	Well, threes and twos.
Res:	threes and twos. …So do you think there might be the same number of threes and twos?
Ivan:	No.
Jim:	Yeah. Well it's going to be more than the fives, fours, sixes, sevens and ones.
Res:	OK.
Ivan:	But it doesn't have to be the same number of threes and twos because it could be just accident.
Res:	So what could we do to be a bit more certain?
Ivan:	Do it again.
Res:	Do another 70 throws? Or another 100 throws?
Ivan:	One hundred.

Ivan stated clearly here that his strategy for understanding what was in the workings box was to collect more data. He clicked the gadget to collect a further 100 outcomes, which were added to the previous 100, and he produced a new pictogram (with a different scale) to display all 200 outcomes (Figure 6). When Jim commented on this graph, he expressed a plan to save this graph, collect a new sample and compare the resulting graph with this one.

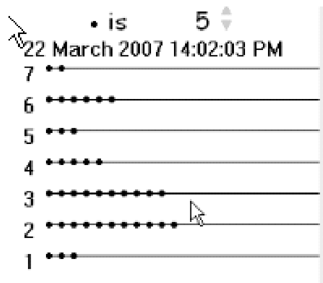

Figure 6. Ivan generated a pictogram for 200 throws.

Jim:	If we kept that, and then made another, made a new one and then did another 4 hundred, we could see if there were actually the same problem. I think that there's more twos than threes cos it says (inaudible)… And we've made it more fours… threes… and that's where we can see which is the biggest.

Jim's proposal for taking a new sample, rather than continuing to grow the existing one, may suggest that, like Rob and Carl, he has recognised that growing the sample is not producing the stability he is looking for. His plan to compare samples may be an expression of a shift in his thinking to a more global perspective. Jim produced the new graph of fifty outcomes and displayed it beside the previous graph of 200 outcomes (Figure 7).

However, as the boys contemplated these images, Jim's attention was again drawn to local changes. He noted that the new graph showed more threes than twos, in contrast to the previous one. The Researcher's questions in response to this remark seemed to prompt

Ivan to look at the data in a new way, as he suddenly began to express a complete description of what was in the workings box.

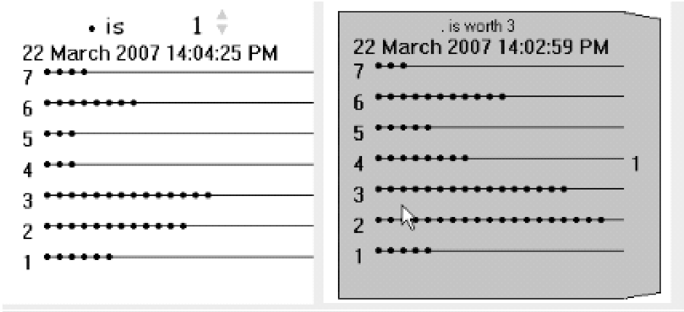

Figure 7. Jim compared a graph of 50 (on the left) outcomes with one of 200 (on the right).

Ivan:	Now there's more threes.
Jim:	…There's more threes than twos. So that means that that could have been randomisation.
Res:	So, do you still think there would be more twos than threes on this dice? Than the other numbers?
Ivan:	Yes.
Res:	What about the other numbers then? Do you think they would all be the same? Do you think that there'd be the same number of one, and four, and five, and six and seven?
Ivan:	I think that there's 2 sixes.
Res:	2 sixes?
Jim:	Well five and four could be the same, but, we did think that two and three would be the same but then two got bigger and then three got bigger, so…
Res:	We've got 12 places to fill up. So there was…?
Jim:	If I divided it by 12…
Ivan:	Yeah, I think… I think I am right…
Res:	Go on then. What do you think?
Ivan:	Cos I think that there's in the workings there's 3 twos, 3 threes and 2 sixes. And that makes 12.
Res:	Ahh. Right. …Say it again, 3…
Ivan:	3 threes, 3 twos and 2 sixes. And that all these numbers for one.

This conjecture was recorded on paper for the boys to consider. Ivan's guess here was in fact correct, but he was not certain of it, and he tried to compare recent graphs with those that they had generated earlier. There was a particular difficulty here as the graphs did not record the number of trials that they showed, and early graphs had been derived from collections of outcomes with differing numbers of trials.

Jim decided to collect 150 more outcomes to add to the 50 shown on the latest graph, and when he did so, the new sample of 200 showed a greater proportion of sixes shown than before.

Ivan noted this and wanted to revise slightly his earlier guess. Faced with the dilemma of which guess was more likely to be correct, both boys suggested collecting some more data. Again they collected a sample of 200 outcomes and again they did not consider that the graph was conclusive. Eventually, the interviewer suggested that they might collect a larger sample of say 500 outcomes. Ivan quickly adopted this suggestion and the resulting graph convinced him that his first guess was correct.

There seems to be evidence, in Ivan and Jim's response to the graph in Figure 7, of a shift towards *perceiving the property* that a large sample size gives useful summary statistics. However, when they looked at the next sample of 200 outcomes (Figure 8), their attention was again drawn to change rather than invariance. Perhaps they were seeking too rigid or robust an invariance, in which case there were aspects of the property still for them to appreciate. An important consideration relating to the property of invariance in distribution is what degree of variability can be accepted while still recognising invariance.

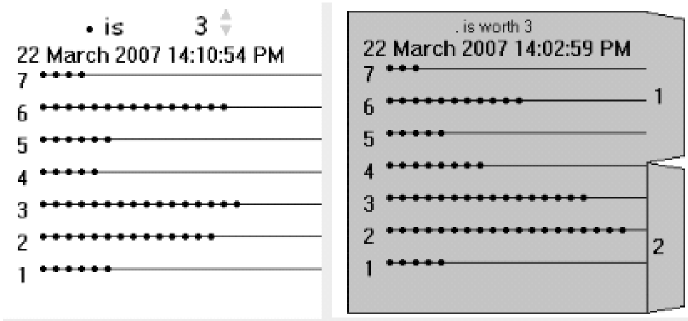

Figure 8. Jim's new sample of 200 throws showed more sixes.

Within our design research approach we adjusted the exact nature of the tasks as we worked with different groups of children in order to explore ways to support their thinking. We tried an approach in which we set a series of challenges for a group of children with dice which were progressively more complex (four sides: [1, 2, 2, 3], six sides: [1, 2, 2, 3, 3, 6], nine sides: [1 2 2 3 3 4 4 4 6]).

Alice, Freya and Bella quickly realised that the four-sided die had only three numbers. After only 12 throws they made a pie chart and recognised that two was occurring more frequently than one or three. Freya expressed this in a proportional way, though it is not clear if she was talking here about the pie chart image, the imagined die, or the workings box.

Freya: The two is double… there's 2 in it … 2 twos

It was relatively rare for children working on this task to use proportional language to describe the images. Generally they talked in terms of relative sizes, but did not attempt to quantify these differences. This might be seen as evidence of their local, rather than global, focus. When we challenged the girls' prediction by showing them a different sample of 10 throws, the image confirmed their decision.

However, Freya's use of language may point towards a power that is released when representations admit the same language for different ways of perceiving; there can also be unintended metonymies, where attention switches between details (such as attributes), because of similarity of language in both perceptions (in this case, the same number). We asked ourselves what Freya was referring to when she said "there's 2 in it." She might be referring to the pie chart image, or the workings box, or the die. Initially, Freya seems to describe the pie chart image ("the two is double"), but, perhaps because there is a potential ambiguity to this reference, she seems to shift towards possibly referring to the workings box or the die. The use of ambiguous language such as this may offer the potential for the speaker (and perhaps the listener) to slide across from referring to the data representation to describing the die, or even the workings box, and in this way to see beyond the data to the sense in which the data provide insight about the population. Such a switch of attention, prompted by the ambiguity in the language used in trying to express what has been observed, might provide a stimulus to 'see beyond the data' and to begin to infer the nature of the population.

The six-sided die proved a little more challenging, but the girls quickly reached a point where they had a number of different conjectures about the distribution of numbers. As they worked, they seemed at times to accept implicitly that doing more throws was a way to test out their conjectures. The following comes from a point where they have done 33 throws.

Alice: one, six, two, three, three, two
Bella: yeah … that's it
Freya: Are you sure?
Bella: Let's do it again, and if the threes keep coming up…
Freya: It's three … it has to be three

The girls continued to add throws without discussing this strategy, and to make further conjectures about the die. They made a new pie chart after each group of 10 throws, and the local changes in this became the focus of their attention. At the stage shown in Figure 9 the three-slice on the pie was larger than any of the others, and the difference between the two-slice and the six-slice was less pronounced. They were struggling to reconcile the image in front of them with what they knew to be possible combinations in the workings box. At one point, they suggested [1 2 2 3 3 3 6], but realised that this would give 7 sides.

Bella: But if it was [1 2 2 3 3 6] I think that the threes and twos would be the same.
Res: Ah!
Freya: Oh yes
…
Bella: The three and the two would be the same size
…
Alice: It's definitely three and two have got the most numbers and one and six have got the least.

63 throws

Figure 9. Alice and Bella draw informal inferences from 63 throws.

Once they had decided confidently that [1 2 2 3 3 6] was their prediction, we challenged this by keeping the pie chart shown above, and making a new sample of 10 throws (Figure 10).

Initially this seemed to shake the girls' confidence in their decision, but Bella soon introduced a different perspective, suggesting that she was considering all the graphs they had looked at, not just those immediately in front of them.

All: It's twos!
Alice: It must be I reckon
Freya: Well it's two and three, two and three
Alice: But look at the six there … it's six and two

Bella: But there's …most of the sixes have been quite low, but most of the threes have been big … if you look at the two it's been right in the middle and it's…

Freya: On that one … on that one it's six and two

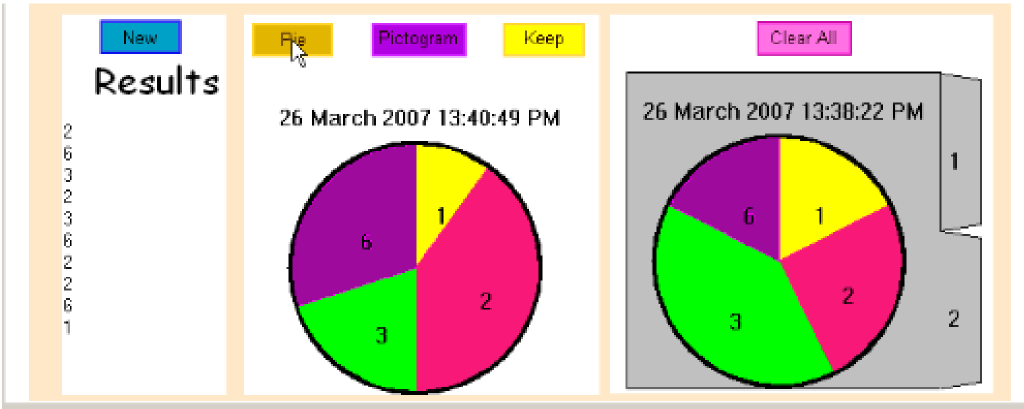

Figure 10. We challenged Bella and Alice with a pie chart based on 10 throws (on the left).

As in the example from Jim discussed earlier, this might suggest that Bella is able at this point to take a more global perspective and see similarities across the set of graphs, rather than focussing on changes between them. Despite Bella's insight, however, the girls continued to be influenced by the new image in front of them, and to make new conjectures about the die. However when asked directly about whether the different number of throws that produced the two graphs might make a difference, the girls agreed that it would, and decided to add more throws to the new sample. They quickly obtained an image which reinforced their confidence in the prediction of [1 2 2 3 3 6].

After their success with the six-sided die, the girls were keen to tackle the next challenge, but found the nine-sided die [1 2 2 3 3 4 4 4 6] much more difficult. Throughout their work they implicitly increased the sample size to get a clearer picture, but actually never went above 100 throws (which they considered to be a very large number). A sample this size had been large enough to show some stability for a six-sided die, but was not sufficiently large to fully explore the nine-sided die. Their strategy was to save graphs, and then begin new samples, but this became complex as they lost track of the sample size for each graph. They continued to struggle between descriptions of the distribution, which seemed to match the images in front of them (but might contain more than 9 numbers), and those which they knew were possible for a nine-sided die, but which did not fit comfortably with the graphs.

6. DISCUSSION

6.1. SAMPLING STRATEGIES IN INFORMAL INFERENCE

When children were working on the task of guessing what the 'funny' die looked like, we saw them use two different strategies which were supported by the software. Having made a graph from a particular number of throws which was inconclusive, they either 'grew' their sample by adding more throws, or they extended the data available to them by beginning a new sample. The latter strategy was supported by a facility in the software to keep a number of graphs, and return to them. In one sense both of these strategies involve increasing the sample size, but the two experiences they provide are very different.

In passing, it is worth mentioning that what is involved in ‘growing the sample’ here is rather different from the activity described by Ben-Zvi and Sharett-Amir (2005), in which children widened the scope of their data collection from a small group of friends, to a whole class, to several classes, and so on. In their activity, it may not be entirely clear, from the children’s perspective, whether it is the sample or the population which is ‘growing’, but the data collected in each iteration can clearly be distinguished. In the case of our task, growing the sample clearly involves collecting ‘more of the same’ data, but the nature of the sample remains the same.

An important question for our thinking about the future development of the task design is how the two different experiences of adding to the sample and taking a new sample impact on children’s local and global thinking. It is clear from our data that neither experience leads easily to the recognition that larger samples are more reliable in providing an image of the distribution; that is, to an appreciation of the Law of Large Numbers. In order to understand this somewhat surprising outcome we need to conjecture about what the children might be *expecting* to see.

We conjectured earlier that Jim was expecting (or hoping) to see invariance in terms of consecutive graphs which stayed much the same as he added to the sample. Given the natural tendency to focus on even small changes, this expectation is almost certain to be confounded, even within sample sizes much greater than those the children were prepared to explore. Seeing repeated change seemed to make some children distrust taking larger samples. Of course taking repeated samples also produced images which reflected change, but possibly the experience of looking at several similar graphs allowed Jim and Bella, if only fleetingly, to gain some sense of the overall stability of the patterns.

However, there is perhaps another way to think about what the children were expecting to see: that is, to see a ‘clear’ pattern. An image in which the sizes of the pie slices, or the lengths of the pictogram bars, were clearly ‘in proportion’ (i.e., some two or three times the size of others) might have proved very convincing, regardless of the size of the sample. Indeed, one group of children did spend time adjusting the scale of the pictogram in order to try to produce such images, with the smallest portions represented by one bead. When children in our study were adding to the sample, they may have been expecting the graph to ‘settle down’ to a clear image, and were frustrated when this did not appear to happen within the sample size they used.

6.2. INFORMAL INFERENCE AS EMERGENCE TOWARDS A TARGET

We are struck by how rarely during the *Guess-my-dice* game in *InferenceMaker* the children referred to luck or chance. Jim was one exception when he was trying to explain to himself the number of observed sixes without abandoning the idea that the die was ‘fair’. It has been well documented how people often do not make sense of phenomena through a stochastic model (as in Konold’s, 1989, outcome approach) or avoid facing the nuances of probability by regarding everything as equally likely, just a matter of chance (as in Lecoutre’s, 1992, equiprobability bias). However, we believe that neither of these interpretations quite fits how the children were trying to infer the nature of the die. We think that these students were trying to see through the data in order to identify the die. Their approach was consistent with Prodromou’s (2007) target connection from the data-centric to the modelling distribution, in which emergence rather than probability is the relevant model.

This is somewhat in contrast to the results observed in Johnston-Wilder's study, where students who were a little older (aged 13 to 18 years) experimented with physical dice; these students therefore did not have such easy access either to larger samples or to graphical summaries of the aggregated data. In Johnston-Wilder's study, the emergence of the modelling distribution was not such a salient feature for the students, and their attention was not so readily drawn to it. Instead, the students were most concerned with judging whether or not the dice that they were using was fair.

6.3. INFORMAL INFERENCE AS THE SEARCH FOR INVARIANCE AMIDST LOCAL CHANGE

Emergent phenomena involve the actions (and often interactions) of many agents at the local level, resulting in the formulation of identifiable patterns at the global level. In order to discern a modelling distribution in *InferenceMaker*, students first need to attend to aggregated data rather than to individual outcomes. They need to focus upon frequencies, and eventually upon relative frequencies, and to pay attention to the pattern of distribution of these across the outcome space, rather than looking only at changes in a single relative frequency from one sample to another. Appropriate graphing tools might support the student in attending to the distribution of relative frequencies. Once a pattern is discerned, and a possible configuration for the die has been conjectured, then this conjecture needs to be tried to see what patterns of outcomes it will produce. The coordination of attention to each of these agents in order to discern the invariance of an underlying emergent distribution, when each manifestation of the distribution in a sample is different, requires several steps, each of which might be supported by developments to the software.

In trying to make a connection from data-centric to modelling distribution, the students needed to identify the trend, a global pattern, that might be emerging from the data. However, our study shows clearly how students' focus of attention, when using *InferenceMaker*, tends to be on the local. Rob and Carl focussed on how the five-slice seemed to be getting bigger, rather than on the dominant size of the six-slice. Jim tended to focus on how there had been more two's but then there were more three's. Alice, Bella and Freya constantly referred to the changes in the new pie chart compared to the previous pie chart. This attention to the here and now, rather than the aggregated longer-term pattern, characterised the activity throughout our trials. Ivan was a clear exception here, who seemed to have a deeper understanding from the start, although even he had required some prompting to consider a sample size of larger than 200 trials. Bella also showed some evidence that she was thinking about the images presented over several graphs. There is plenty of evidence that the students wanted to find invariance but were constantly frustrated because all they could see was change. In Mason's terms, the students were unable to hold the same wholes that more experienced statisticians might do because the properties of invariance were constantly hidden by the tendency to perceive change. The task of identifying the invariant properties was made even more difficult because invariance in the aggregated whole is not absolute invariance but relative invariance; to identify the property of relative invariance, one has to notice how the changes in proportion becomes less significant as the sample gets larger.

6.4. INFORMAL INFERENCES ARE OFTEN MADE ON SMALL AMOUNTS OF DATA

We are struck by an apparent paradox that students wanted to generate more and more data, as they were slow to feel confident about their conclusions, and yet they placed no greater confidence on inferences based on large amounts of data than those made from small amounts of data. Konold (1995) has in the past made the point that data are not forceful in persuading people, though our students did seem to choose to generate more data.

We believe this paradox might be resolved from the perspective of emergence. The students were looking for stability at the wrong level. They hoped, in their search for invariance, that by collecting another batch of data, they would get the same pie chart or pictogram as the last one. We have discussed above how different strategies for collecting the additional data shaped the attention.

Inevitably, whichever way they sought to collect data, there would be change which would attract attention away from their search for invariance at the global level towards the local level where invariance could not be found. How that data were displayed could also make a difference. Pictograms tended to emphasise differences between the lengths of bars in the case of dice with fairly uniform distributions but would be effective in showing up patterns in more distinctive distributions (such as would be the case with the frequencies of the totals of two dice). Pie charts tended to be ineffective in showing up such patterns but reduced the apparent differences between sectors.

Nevertheless, even when using pie charts there was a powerful attraction towards local change. And so, in practice, the extra data provided only more complexity. Carl pointed out that more throws were “just getting too stupid” because they “had too much in.” Somehow, Carl needed to focus at the global level to find the stability he wanted.

Eventually, the students would sometimes see no added value in generating yet more data and would be content to make their inference on what we might regard as flimsy evidence.

7. CONCLUSION

When we argue that young students’ naïve informal inference is emergence-related (rather than based on chance), focussed on the local, and made on small amounts of data, we do not present these as misconceptions to be eradicated. Rather, we see these findings as identifying students’ starting points, informing how we should be building new designs for tasks and learning environments which will offer to students experiences that may enable them to construct more sophisticated meanings for informal inference out of these relatively naïve conceptions. And we do not see this as a forlorn hope.

We believe that, although inference involves making a connection from the data-centric distribution to the modelling distribution, this connection is supported by the intention connection in the opposite direction. We conjecture that giving students the experience of mending gadgets before asking them to infer the nature of the die may be a necessary experience to enable students to understand more deeply the connection from data to modelling distribution. Seeing that random mechanisms can generate many different looking pie charts when the data are limited may be another vital experience. For comparison, perhaps the students needed to see the stable patterns generated by a large number of throws. At the outset, we worried that the students would simply throw the die 1000 times and immediately infer the nature of the die. However, this simple strategy was not available to them because they were not paying attention to the global level of emergence. Indeed, it is likely that they had no global resource such as the Law of Large Numbers available to them.

But suppose they did have. Would this mean that the students were in a powerful position to make informal inferences? Possibly not. Then, it would be interesting to explore the level of confidence students place in their inference if they are allowed only a limited number of throws. How would we support changes in thinking towards an abstraction, which might be schematised as "the more data we have, the more confident we can be in our inferences"?

In considering a new design for *InferenceMaker*, we recognise the key objective is to support students in their attempts to observe global relative invariance in the midst of local change. Such an aspiration leads us to consider (i) increased support for systematic recording and reflection, and (ii) functionality for exploring the behaviour of conjectured die configuration for comparison with the behaviour of the unknown die. More specifically

- We recognise the need to enable systematic recording of the students' conjectures about what the die looks like at any particular time through entering the conjectured sides into a blank 'die' with the correct number of sides.
- We would then consider allowing experimentation with the conjectured die for comparison with what the unknown die has generated. In effect, the students would be creating a workings box and generating results from it, thus enabling the possibility of intention connections from the modelling perspective to the data-centric perspective without losing contact with the challenge of finding the configuration of the unknown die.
- We would consider the provision of graphing and summative tools to enable students to compare the conjectured die with the unknown die. (For example, we can imagine deploying graphing and modelling tools of the type available and becoming available in Tinkerplots™, Konold & Miller, 2001.)
- Finally, we believe there would be value in enabling better recording of the number of throws with easy cross-reference to the display generated.

From the research perspective, it would be valuable to register what appears to be the focus of the students' attention, local data or global effects. We would therefore want to place more attention on recording aspects of body language, gesture, and where students appear to be looking to augment the Camtasia records.

We reconsider the differences described at the start of this paper between the perspectives of local and global thinking presented in the previous studies of Pratt and Johnston-Wilder. The theoretical lens of Structure of Attention may allow us to see these as products of the available technology and task design, rather than as more fundamental discrepancies. Pratt's students, unlike Johnston-Wilder's, were not able to focus on the sequential outcomes of throwing the dice. Although this is available on the screen, the layout makes it difficult to see, and when using the facility to generate data quickly through groups of trials, the individual results appear too quickly to allow attention to focus on each individually. This might be seen as an advantage, because students are unlikely to become distracted by local sequential patterns, but it may also disguise a potential opportunity to move between local and global perspectives. Similarly, Johnston-Wilder's students did not have access to the graphing facilities provided in Pratt's software, which, although they provide tools, may also serve to disguise the growth of the sample, because each graph may appear to be 'the same size'.

As a final reflection, we return to Makar and Rubin's analysis (2007) in which they identified aggregate thinking and sample size as two important components of informal inference. Through the lens of Structure of Attention, we have seen how our students, aged 10-11 years, were drawn to local variation and often the invariant characteristic of relative frequency, apparent in aggregate thinking, was obscure to them. By being aware of the focus of the students' attention, we have not only begun to appreciate why inference may be such a problematic area but also how the design challenge should begin to respond.

REFERENCES

Ben-Zvi, D., & Sharett-Amir, Y. (2005). How do primary school students begin to reason about distributions? In K. Makar (Ed.), *Reasoning about Distribution: A collection of studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4).* [CDROM, with video segments]. Brisbane, Australia: University of Queensland.

Ben-Zvi, D. (2006). Using Tinkerplots to scaffold informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute. [Online: http://www.stat.auckland.ac.nz/~iase/publications/17/2D1_BENZ.pdf]

Camtasia Studio (Version 6.0) [Computer software]. Okemos, MI: Techsmith Corporation. [Online: http://www.techsmith.com/camtasia.asp]

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

Johnston-Wilder, P. (2006). *Learners' shifting perceptions of randomness*. Unpublished doctoral dissertation, Open University, UK.

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

Konold, C. (1995). Confessions of a coin flipper and would-be instructor. *The American Statistician*, *49*(2), 203-209.

Lecoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics*, *23*, 589-593.

Makar, K., & Rubin, A. (2007, August). *Beyond the bar graph: Teaching informal statistical inference in primary school.* Paper presented at the Fifth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-5), University of Warwick, UK.

Mason, J., & Johnston-Wilder, S. (2004). *Fundamental constructs in mathematics education*. London: RoutledgeFalmer.

Noss, R., & Hoyles, C. (1996). *Windows on mathematical meanings: Learning cultures and computers*. London: Kluwer Academic Publishers.

Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute. [Online: http://www.stat.auckland.ac.nz/~iase/publications/17/6A2_PFAN.pdf]

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. (2007). Making connections between the two perspectives on distribution. In D. Pitta-Pantazi & G. Philippou (Eds.), *Proceedings of the Fifth Conference of the*

European Society for Research in Mathematics Education (pp. 801-810). Larnaca, Cyprus: University of Cyprus.

Prodromou, T. (2008). *Connecting thinking about distribution*. Unpublished Doctoral Dissertation, University of Warwick, UK

Prodromou, T., & Pratt, D. (2006). The role of causality in the coordination of two perspectives on distribution within a virtual simulation. *Statistics Education Research Journal, 5*(2), 69-88.
[Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ5(2)_Prod_Pratt.pdf]

Smith, J. P., diSessa, A. A., & Rochelle, J. (1993). Misconceptions reconceived - A constructivist analysis of knowledge in transition. *Journal of Learning Sciences, 3*(2), 115-163.

Konold, C., & Miller, C. (2001). Tinkerplots (version 0.23) [Data Analysis Software] University of Massachusetts, Amherst (USA).
[Online: http://www.keypress.com/x5715.xml]

DAVE PRATT
Institute of Education
University of London
20 Bedford Way
London
WC1H 0AL