WORKING GROUP 5. Stochastic Thinking 685

DEVELOPING STOCHASTIC THINKING – A WORKING GROUP REPORT OF CERME 5

Rolf Biehler, Universität Kassel, Germany

and

Maria Meletiou, Cyprus College, Cyprus
Maria Gabriella Ottaviani, "Sapienza" University of Rome, Italy
Dave Pratt, University of London, England

1. OVERVIEW

The working group discussed 15 papers on various aspects of stochastic thinking. We organized the papers into 4 sessions with the following titles:

- The relationship between stochastic thinking & knowledge and external factors such as teaching methodologies, tools, tasks and setting
- The role of computer-based tools, including microworlds, on stochastic thinking/knowledge
- Research in primary school or preschool
- Research on students' understanding of different concepts

The structure was a pragmatic one that tried to establish adequate discussion contexts between the papers. The papers often touched several dimensions such as when papers on teaching in primary school were also concerned with the use of technology. We will use a structure that is a little different from the original one for reporting on the session in order to enlighten four fundamental factors in educational context: teachers, curriculum, students and information technology.

2. TEACHERS' KNOWLEDGE AND BELIEFS

Teachers' knowledge and beliefs on statistics and probability influence how and what they teach in the classroom. Teachers are the most important agents of any educational reform. Andreas Eichler's paper focuses on a research project that combined three aspects of a curriculum namely teachers' planning, teachers' classroom practice, and their students' statistical knowledge. The planning and classroom practice of one "expert" statistics teacher (upper secondary level) was analyzed, as well as the knowledge and beliefs of some of his students by means of interviews. Findings indicate that this expert teacher had already developed a very detailed and broad individual view on stochastics during his yearlong teaching experience. This influenced and constrained the beliefs, as well as the content and structure of knowledge that students could develop in his classroom. On the other hand, the study showed how different students' knowledge might be from the knowledge the teacher intends to impart. Educational change has to take this situation into account. Expert teachers' knowledge and beliefs are difficult to change, and they may reflect boundary conditions of teaching at school level, which also have to be taken into consideration.

Efi Paparistodemou also studied (pre-primary) prospective teachers' views of randomness as they are expressed in the tasks given to the students. We discussed the important role of pedagogical content knowledge that can bridge pedagogical and content related aspects of knowledge for prospective teachers.

2. CURRICULAR INNOVATION AND CURRICULAR CONSTRAINTS

The teacher is the most important mediator of the intended curriculum. Several papers also touched the question of what should be the content of stochastics curricula. These include most papers that are concerned with the use of technology and that considered new content besides discussing the role of new IT-media as a teaching tool, which generally are not neutral to the question of content.

The paper by Marta Carles Fariña and Pedro Huerta is concerned with conditional probability and Bayes' theorem as content. A systematic analysis was done to reconstruct the reference knowledge of this domain on which curricular decisions at the school level should be based. University text books and uses in various application contexts were regarded as a basis for constructing this reference knowledge. The results of the study showed the limited problem types and contexts that are currently in use at the school level in Spain.

The discussion raised awareness of the different traditions in different European countries with regards to what is research in didactics of mathematics. In the tradition of Yves Chevallard's "transposition didactique", the construction of curricula has also to be based on scientific research on what is the knowledge to be taught. In other scientific traditions, this is not seen as an essential prerequisite.

The discussions made us aware of how the different studies that were conducted on students' understanding are, at least implicitly, based on certain implicit views of stochastics. The multinational group discussions were very good for coming to an increased awareness of this problem. Studies have to be judged differently based on whether they try to assess national curricula, in which case the test has to be curricular valid, or whether they start from a normative conception of mathematical and statistical knowledge that is to be assessed. This is the approach PISA adopted.

The need for large scale tests that reflect the state of the art with regard to statistical literacy, thinking and reasoning was felt, however such studies cost a lot of money and state authorities as clients may have specific different interests.

We discussed large scale studies with regard to their implicit conception of stochastical knowledge. Maria Meletiou & Carl Lee's paper is concerned with students' knowledge of graphs prior to an introductory statistics course. There was also an interesting discussion on assessing probability and statistics in the British National curriculum stimulated by input from Thekla Afantiti and Maria Pampaka.

A specific concern of the group was curricular innovation at the preschool, the primary and the lower secondary level. The papers discussed background research, conceptions, tools and working environments for students. Some studies were done at first in a laboratory context, where experimental conditions can be better controlled. The group agreed that long-term studies of innovations that Michele Cerulli and his group did are a very important next step, if we think of implementations of new tools and conceptions under normal classroom conditions. An innovation has to take the whole system into account, not only a new tool or conception, but also teacher education and development of the classroom culture.

Laura Martignon's paper provided a profound review of research in cognitive psychology concerning natural frequency representation formats that foster the development of stochastic thinking in relation to proportional thinking. Based on this research, new enactive working environments with tinker-cubes and transparent urns have been designed for primary education and tested in exploratory studies. Marta Carles Farina, Ma. Angeles Lonjedo and Pedro Huerta also discussed the topic of natural frequencies as a superior representational format in their studies with regard to the secondary level.

The role of dynamic representation was also central in the computer-based environments that Michele Cerulli and Theodosia Prodromou studied in their papers. The co-ordination of two perspectives of distributions (frequentist and theoretical-combinatorial) is a basic problem for young children. Simulations with adequately designed representations and interactional options for the children can help to build understanding. Conceptually this is also related to the problem of recognizing equivalent sample spaces, and of building bridges between probabilistic and causal-deterministic thinking.

Equivalence of sample spaces and its relation to outcomes of random experiments was also a central issue in the study of Zoi Nikiforidou and Jenny Pange with preschoolers.

All studies suggest curricular innovations. Some of them still intend and need further careful testing under laboratory conditions.

Curricular innovation using simulations was also discussed with regard to secondary and tertiary level with regard to topics such as modelling in elementary probability (Carmen Maxara & Rolf Biehler) and confidence intervals (Hermann Callaert). There was also an interesting discussion on using simulation to support resampling approaches to statistics stimulated by input from Manfred Borovcnik. In addition to using simulation, very careful considerations about the symbolic representations and tools used together with simulation were discussed.

3. STUDENTS' COMPETENCIES AND UNDERSTANDING

Students' understanding of specific domains of probability and statistics were discussed. This was partly done in the context of studying the effect of innovations and interventions, partly as a general assessment of certain representative groups of students, partly related to theory driven experimental work under laboratory conditions. The studies were mostly related to topics that we also described in the preceding two sections.

The topics include

- Studies on probability knowledge
    - relation between probability models, causal ideas, and experimental data with random devices (local and global meanings, informal inference)
    - events, sample spaces and random variables
    - conditional probability, Bayes'reasoning (different representation formats, natural frequencies)
    - simulation and resampling
    - confidence intervals
- Studies on statistical knowledge
    - Graphs
    - Measures of center and spread

Progress has been made on different approaches of identifying (conceptualising) and measuring students' competencies; approaches are still very different and not yet related to a shared body of research knowledge and methodological standards. Different approaches include

- Hermeneutics and case studies
- Rasch scaled large written test

A topic that reoccurred in many of our studies was the first topic of the relation between probability models, causal ideas, and experimental data with random devices (local and global meanings, informal inference). This includes studies on the use of simulated data from which students draw inferences to probabilities. Peter Johnston-Wilder & Dave Pratt were concerned with how students develop local and global meaning of randomness, a study that was related to Theodosia Prodromou's study on two notions of distribution, as well as to various other papers which discuss ways in which frequency representation formats can support the development of students' stochastic thinking.

The concept of probability is inherently very complex and very different from other mathematical concepts. Even if we take a simple device such as a die, we can determine probabilities by the classical approach as the quotient of the number of favourable outcomes to all possible outcomes. But we can also experiment and estimate probabilities by frequencies. These approaches have to be co-ordinated. The sequence of results is unpredictable but nevertheless displays certain patterns, and in the long run the relative frequencies stabilize. Moreover, the die is a physical deterministic device, but it can be treated as random at the same time. This ambiguity is difficult to understand even for adults. This complexity is one of the reasons for the large number of studies addressing aspects of this problem.

Some of the studies constructed special artificial worlds where students can experiment with idealized devices that are partly deterministic and partly stochastic, which is very difficult for instance to do in real situations such as changing the shape of a die or systematically varying the starting conditions of a ball whose movement is a mixture of the laws of physics and random noise.

4. THE ROLE OF COMPUTER-BASED TOOLS, INCLUDING MICROWORLDS, ON STOCHASTIC THINKING & KNOWLEDGE

The group discussed the implementation of various tools in a range of environments (software, task, settings…)

- Tools: Fathom, Excel simulations
- Environments: Toontalk based microworlds; Logo-based simulations; the ChanceMaker simulation;

A range of qualitative methods were employed:

- Recording of students' and screen interactions
- Whole classroom approaches

We have worked on the task of relating the world of technology with the mathematical concepts and notations and generated new unresearched ideas about using technology rooted in a priori didactical analysis.

As was mentioned in earlier sections, the experimental environments were constructed and studied with regard to the knowledge and conceptions that students develop when experimenting with the environments. The interface design was made as easy as possible so that young students can work with the technology using simple metaphors. The tool complexity and the difference between conceptions and technological implementations should be as narrow as possible.

In contrast to this, studies related to complex tools such as EXCEL and FATHOM for secondary and tertiary students, were interested in the process of *instrumental genesis* through which the students begin to use the artefacts as instruments for solving problems. This process has to be supported by adequate measures. Carmen Maxara & Rolf Biehler report on using "simulation plans" for helping students to use FATHOM in modelling and simulation problems. In this context, it is not sufficient to only study the conceptual development of students, but also to examine how they use and learn to use the tools as an artefact. If we pay attention to the micro processes involved when students collaborate on using the technology, we can gain insight on these processes.

This is one of the reasons why qualitative methodologies dominated in these studies as compared to quantitative comparative studies, where the global effect of one set of innovations is compared to traditional teaching.

5. CONCLUSIONS

The group work was seen as very useful and efficient. In contrast to specializing sessions in large conferences, the papers in our small group were much more diverse in content, school level and methodological approach. Having the possibility to work over a long time period provided the opportunity for very intensive discussions on topics and research approaches that need not be very similar to each other but that may very much profit from the expertise of others working in different domains of stochastics education. We had introduced two “reactors” for each paper that we discussed. The reactors had the opportunity to prepare 5-minute statements on the respective paper. This turned out to be a very effective way to stimulate discussion, to get many participants involved, and to insure high quality papers to be published.

UNDERSTANDING CONFIDENCE INTERVALS

Herman Callaert

Center for Statistics, Hasselt University, Belgium

herman.callaert@uhasselt.be

This paper addresses statistical reasoning at the level of a first course in statistics in higher education. It focuses on two major difficulties encountered by students when studying confidence intervals. The first difficulty relates to the randomness of intervals in repeated samples. Students' understanding might be enhanced here by learning how to think backwards, the use of computer-supported tools and adherence to a specific notation. The second difficulty relates to the non-randomness of a confidence interval after the sample has been taken. Circumstantial evidence shows that a simple story about a Wheel of Fortune can vastly improve students' insight. It might be worthwhile to investigate this evidence further and to find out whether this illuminating story can successfully be translated into a concrete tool.

INTRODUCTION

As a teacher of statistics, there is plenty of evidence from my own experience and from the experience of many colleagues, that the proper understanding of confidence intervals seems to be unusually difficult. This difficulty stems from two sources, namely the understanding of the variability of models and the understanding of the non-randomness of a sample result.

A lot of classroom material (and several research papers) focus on the variability of the intervals in repeated sampling. The understanding of this variability is a major problem to many students and it relates to the observation made by R. Biehler (1997b) who writes: "… working with functions as entities is difficult for students… This difficulty may not be surprising because data distributions are usually not characterized as concepts in courses of elementary data analysis. Distributions are emphasized in probability theory but in an entirely different context that students find difficult to apply to data analysis."

Apart from the difficulty of thinking in global models there is another stumbling block that comes with confidence intervals. The problem has to do with the non-randomness of the confidence interval after the sample has been taken. At that point in the statistical study (and in the formulation of the conclusion) there is nothing random any more. This fact is mentioned in many textbooks but rarely exemplified with concrete stories or tools. Mistakes appear frequently here, and students seem to need concrete examples they can easily and quickly relate to. Anecdotal evidence suggests that a story about a Wheel of Fortune might help in students' understanding of the second part in the thought process surrounding confidence intervals. It might be of interest to further explore this evidence through systematic and formal research.

A related question might be about the age and the level of maturity at which students are able to grasp this type of reasoning. Further research could be carried out for investigating to what extent our observations can be transferred in different settings, real or in mind, and in particular in a computational environment.

THINKING BACKWARDS

If you tell students that you are going to throw a honest die and you ask them what the result will be, their immediate answer is: "how can I know?". It takes some time for discovering that they nevertheless know a lot about the outcome of a die. It will never be a seven. Indeed, they know what the possible outcomes are and even with what probability they will occur. This is a completely different way of thinking about the outcome of a die. It has to do with the probability model as an ideal mathematical construct for modelling outcomes in a physical world. If this is the kind of probabilistic thinking needed for a proper understanding of statistical inference, then we fully agree with Cobb and Moore (1997) who write: "What then, should be the place of probability in beginning instruction in statistics? Our position is not standard, though it is gaining adherents: first courses in statistics should contain essentially no formal probability theory. …. The concepts of statistical inference, starting with sampling distributions, are of course also quite tough. We ought to concentrate our attention, and ours students' limited patience with hard ideas, on the essential ideas of statistics".

Research by Pratt (1998, 2000) nicely illustrates how children behave differently with respect to the interpretation of unexpected "outcomes" and with respect to their ability to manipulate "the underlying chance mechanism" (workings box). The insight that random outcomes can be modelled by a chance mechanism generating them, is an essential idea in statistics.

At the time students enrol for a first course in statistics, they should be able to think backwards when observing random outcomes. They should have learned to always search for an underlying chance mechanism while realizing that the random outcomes do not perfectly coincide with the proposed model. A simple example is as follows. A physically honest die is thrown 450 times and one observes 148 ones, 155 twos, 69 threes and 78 fours. If you know that Ann was allowed to write any number on any side of the die, as long as she choose from {1, 2, 3, 4, 5, 6} (repetitions allowed), what did she do? Remark that the observed proportions can't coincide with the underlying chance mechanism if it has to be a die.

Students who are familiar with some formal mathematical notation could benefit, in a first course on statistical concepts, from a systematic use of a specific notation. Population parameters like the population mean or proportion are denoted by the Greek letters μ and π. Outcomes, based on observed results in a sample, are small letters. A universal convention is $\bar{x}$ for the sample mean and one often encounters $\hat{p}$ for the sample proportion. At this point again, the difference between an outcome and the underlying chance mechanism is crucial and should be made clear through a distinct notation. In analogy to the use of capital letters for representing the chance model of a random variable, it is proposed to use capital letters for the chance mechanisms of outcomes based on samples. The underlying model for the sample mean is $\bar{X}$ and $\hat{P}$ is the notation for the chance mechanism of the sample proportion. This convention is systematically ignored in almost all major textbooks in the United States. As examples, the books by Agresti & Franklin (2007), Yates, Moore & Starnes (2003), and Watkins, Scheaffer & Cobb (2004) all use capital letters for random variables but small letters in their discussion of "the behaviour" (often called "sampling distribution") of the sample mean or the sample proportion. It might be an interesting research question to find out whether the use of a separate notation for underlying chance mechanisms helps in the better understanding of statistical ideas for the mathematical capable student.

THE CHANCE MECHANISM OF RANDOM INTERVALS

After realizing that random outcomes are generated by an underlying chance mechanisms, a further step is the observation that "an outcome" can be "an interval". Understanding the underlying chance mechanism is a first step in the study of confidence intervals.

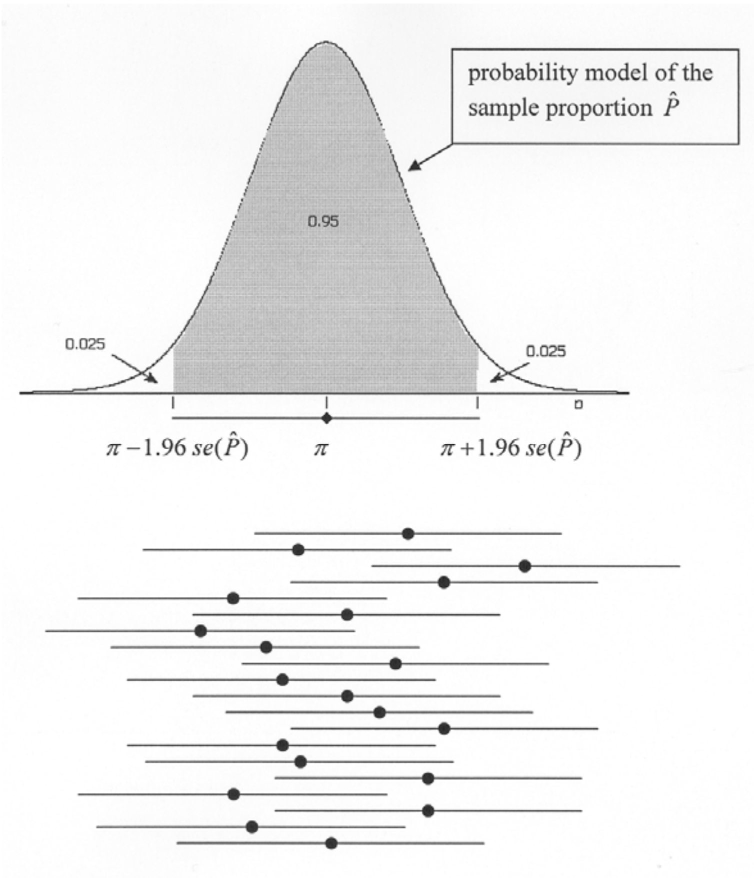


Figure 1

Sometimes the explicit underlying model is not formally introduced but instead the variability of the calculated interval in repeated samples is illustrated through pictures and simulation (see e.g. Biehler (1997a) p. 184). Attention is drawn on the fact that, on average, 95 intervals out of 100 will capture the population parameter. If the research question is about a population proportion π, a classical illustration is as in figure 1. This figure illustrates the behaviour of the underlying chance mechanism for generating intervals when the sample size is large and the normal approximation can be used. However, that figure might divert students' attention from the fact that in their statistical study they will end up with just one out of all those pictured intervals.

A STUMBLING BLOCK

The difference between an underlying chance mechanism and a particular outcome seems hard to grasp in a somewhat complicated setting of the construction of confidence intervals. Sometimes students get away when they are allowed to write down conclusions "in words" where they can "talk around" the problem. But the more mathematically sophisticated students should also be asked to formulate their result in mathematical notation. One then encounters a variety of expressions where sample values are plugged in while keeping a probability statement about fixed numbers as seen in figure 2.

$$P\left\{\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right\} = 95\%$$

For this case: $P\{403.82 \leq \mu \leq 406.18\} = 95\%$

Figure 2

Another frequent type of error shows up when the student "mechanically" knows that he should switch from probability to confidence but where he keeps writing inequalities about a fixed but unknown population proportion π as in figure 3.

The actual 95 % Confidence Interval is

$[12,7\% \leq \pi \leq 18.9\%]$

Figure 3

These (and similar) types of error keep showing up over and over again. I have encountered them every year in my long experience as a teacher of statistics.

THE WHEEL OF FORTUNE

Anecdotal evidence suggests that the wheel of fortune (or a similar device) might be a crucial tool in student's better understanding of what is really going on with confidence intervals. It might be of interest to carry out more extensive and more formal research on the particular way in which this didactical tool is used in this context.

Contrary to devices that help in the understanding of sampling variability, the Wheel of Fortune confronts the student with the second step in the reasoning about confidence intervals. It is the step where the student has to formulate a conclusion after the sample is taken and after the confidence interval is constructed. In order to let the basic problem stand out clearly, no attempt is made here to also incorporate the (important) interplay between confidence, precision and sample size.

A basic difference between a spinner and a Wheel of Fortune might lie in the purpose for using the device and especially in the emotional expectation about the outcome. A spinner focuses on (a.o.) randomness and on underlying models for random outcomes. On the other hand, when playing the Wheel of Fortune in a television show, people are not interested in long run properties of random outcomes. In a simplified version, their only question is: will the wheel stop and give me the ticket with the jackpot or will I be broke? The main emotional stress and attention goes to the single sector on the wheel (or the single ticket) where the wheel stops and the game is over.

Let's try to illustrate this concept assuming that a 95 % confidence interval is needed for a population proportion π. Let's also assume that the student knows how to set up the appropriate model. He also understands that the use of this model assures, as a long run property, that on average 95 out of 100 intervals will contain the population proportion π. He learned this property through simulations with a spinner-like device.

He now has to play the Wheel of Fortune in a somewhat modified version. In this game he has several aspects under his own control. He can decide in how many sectors of equal size the wheel has to be partitioned. Since he wants to work with a procedure that produces "good" intervals (containing π) with probability 95 % he asks for a division of the wheel into 100 sectors. If he would have decided to work with a 99.5 % confidence interval he then might have asked for 1000 sectors.

The student then tells the quizmaster to firmly stick tickets on each of those 100 sectors. On each ticket an interval of the form [a ; b] is written with a and b both numbers between zero and one. The student also tells the quizmaster that exactly 95 of those tickets should have a "good" interval written on them. The quizmaster, who knows the value of π, does exactly what is asked for. He doesn't tell which tickets are the good ones. Then he leaves… forever.

Now it's the student's turn. He spins the wheel and after some time it stops at one sector. The student has to take the ticket on that sector. With that ticket in his hand he now has his interval, with the fixed endpoints written on. Nothing is stochastic anymore. It could be [0.26 ; 0.32] like illustrated in figure 4.

On the other hand, the student doesn't know the value of π and he cannot check whether his interval is a good one since the quizmaster is gone. But the student knows that there is nothing stochastic about π either. It is a population parameter, unknown but fixed.

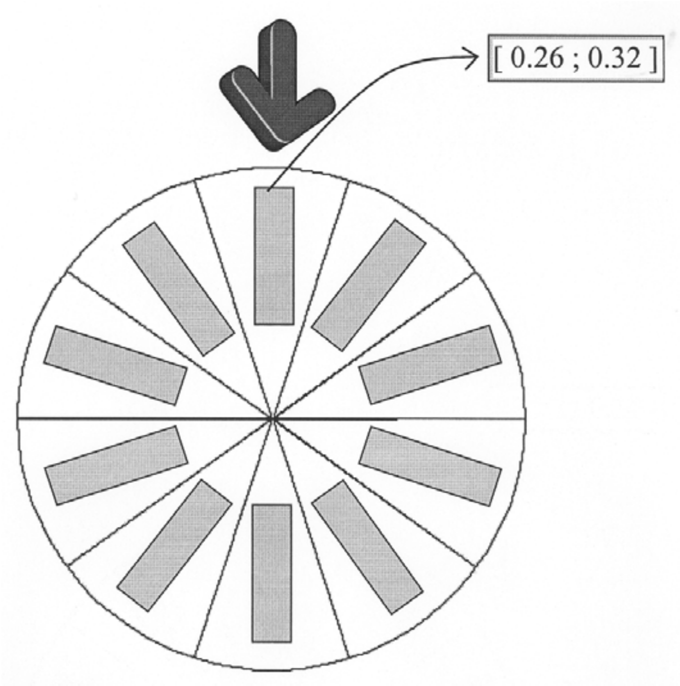


Figure 4

Combining these two deterministic facts leads to the straightforward conclusion that the student can be 100 % sure that his 95 % confidence interval either contains π or doesn't contain π, and that he will never know. Stating the conclusion in such a blunt way is shocking to most students. It makes them think all over again, since this can't be the purpose of constructing confidence intervals… or is it?

NEVER CERTAIN, SO WHAT?

The shock effect of the perplexing conclusion usually generates a beneficial reaction with many students. They now go and search for a deeper understanding of what they thought they already understood. They are forced to think backwards, since that one ticket in their hand with that one confidence interval can't be the whole story.

At this point, many students succeed in putting their ticket into the right context. They got it through a procedure where 95 tickets out of 100 were good ones. That means that before spinning the wheel there was a probability of 0.95 for getting a good ticket. After spinning the wheel, they hope for the best but they never can be sure. Realizing that this is the way they have to think about statistical inference is a big step forward in a deep understanding of statistical reasoning.

CONFIDENCE IS …

The word confidence, as used in the context of confidence intervals, is strange and unnatural in Flemish, my native language. I assume that "I am 95 % confident" doesn't belong to the daily vocabulary in English either. The same may be true in many other languages. Reading copies of AP exams (USA), the uniformity in the formulation of answers on confidence intervals is striking. Every student with a correct answer uses the above sentence, exactly as it stands there. Nobody dares an explanation "in his or hers own words". Of course, those students have learned that the word "certainty" is not a good choice when talking about the interval catching the population proportion π. They also have learned that the word "chance" is not a good choice, neither in combination with the parameter π nor in combination with their calculated interval. So, there apparently is nothing left but "I am 95 % confident", whatever that means.

The problem is real, and it has probably nothing to do with language. The main point is that "95 % confidence" refers simultaneously to two crucial but quite different steps in the construction of confidence intervals. Reference to the Wheel of Fortune can be revealing here.

The "95 %" refers to the underlying model for generating intervals or to the Wheel of Fortune where, before spinning the wheel, it was under the student's control to decide how many good and how many bad tickets there should be. In this example, the student asked for 95 good ones out of 100 and the 95 % refers really to a probability of getting a good interval when you will spin that wheel.

The word "confidence", and not probability, refers to the fact that you already have played the game. After spinning the wheel you hold that single ticket in your hand. It can be a good one, it can be a bad one.

"How should we proceed now?" is the common further question. The answer again reflects some deep statistical ideas. Indeed, the only information available comes from the sample together with the insight in how the confidence interval came about. So, for the rest of the study you can work with a value of π that belongs to the values of your confidence interval. They are for you "plausible" values, based on your sample. You do not have anything better. You do not have certainty either.

SOME FURTHER POINTS OF REFLECTION

To start with, it should be repeated that this paper is about confidence intervals as they are usually taught in a first course in higher education, in a classical frequentist framework, where a population parameter does not have a (Bayesian) distribution associated with it but is a fixed number. The fact that the value of a population parameter is unknown doesn't make it random.

The experience of statistics educators

An interesting source of information is the AP Statistics EDG (Electronic Discussion Group) which is a monitored electronic discussion list accessible through http://apcentral.collegeboard.com/apc/public/homepage/7173.html . Teachers of AP statistics courses discuss the many problems encountered in class and the discussion is often joined by university professors in statistics. An AP statistics course is a course at the level of a first introductory statistics course in college (university) but the course is taught in high schools by high school teachers. Each year in May there is a central nationwide (USA) exam and an excellent performance by the student may lead to a credit for statistics when he/she enters college (university) in September.

Explaining confidence intervals seems extremely difficult as exemplified in the many discussions during Januari-March 2007 on the AP Statistics EDG. Main stumbling blocks are illustrated in the following excerpts.

Dave Bock (Jan 29 2007) points to the difference between “the underlying process” and “one particular realisation that has happened”. He gives precise rules to the students about what they “may” and “may not” say on the exam.
"Things you MAY say:
(1) I'm 90% confident the population mean is between 20 and 30. (Interprets the interval)
(2) If this study were conducted many, many times, I expect that 90% of all the intervals created from the various samples would contain the population mean. (Interprets the confidence level)
Things you may NOT say:
(3) The population mean is between 20 and 30 90% of the time. (Either it is in this particular interval or it isn't.)
(4) The probability the population mean is between 20 and 30 is 90%. (Slightly more sophisticated-sounding restatement of 3)”

Daniel J. Teague (March 11 2007) warns about the use of “I am confident..”.
I strongly discourage you from letting your students say, "we are 95% confident in ..." I know that will be counted correct on the exam, but I simply don't believe kids can say "95% confident" without thinking "probability of 0.95".

Probability or confidence or what?

Students have difficulty with the interpretation of a confidence interval.

Kenneth D. Nilsen (AP Stat EDG, Feb 3 2007) states the problem very clearly.
If I'm asked to "interpret" the confidence interval, and I say "I'm 95% confident that the true average is between 575 grams and 710 grams", I don't think I have "interpreted" anything.

The answer by Robert W. Hayden (Feb 3, 2007) might seem shocking, but it is certainly recognizable by many statistics teachers and students.

To me the "95% confident" phrase is one of those compromises that grows up between student and teacher.
• Teachers like it because it's easy to present and isn't really wrong;
• students like it because it provides a nice mantra that the teacher will accept without any requirement for the student to know what's going on…..
I agree with you 100% that it is NOT an "interpretation" -- it is a mantra, in this case,…. a verbal veneer to hide lack of understanding.

Daniel J. Teague (March 11 2007) warns about the use of the word "confident".
I strongly discourage you from letting your students say, "we are 95% confident ..." I know that will be counted correct on the exam, but I simply don't believe kids can say "95% confident" without thinking "probability of 0.95".

A number doesn't turn into a random variable just because you close your eyes

Probability statements are about outcomes to be generated by a random process, not about a particular realisation of this process after it has happened. There are several ways teachers try to explain this difference.

Kate Baker (AP Stat EDG, March 12 2007) starts with simple chance processes before turning to confidence intervals. Remark the difference between "*probability of getting*.." (which refers to the process) and "*probability I got*" (which refers to the particular realisation after performing the experiment).
... Holding up deck of cards: What is the probability of getting a queen, if I draw card from the deck? Draw top card, the queen, show the class: What is the probability I got a queen? Draw next card, the not queen, show the class: What is the probability I got a queen?
Draw next card, don't show class: What is the probability I got a queen? We discuss how it is the same as the previous two draws, 0 or 1, we just don't know. The probability is associated with the method of drawing cards, the experiment, not the result.
I make them take notes on this and refer back to it when introducing confidence intervals. At the time, they thought I was an idiot. But when we got to CI, they really seemed to get the idea better than ever before.

CONFIDENCE INTERVALS AND TESTS OF HYPOTHESIS

The main problem in the exact formulation of the result after the sample has been taken is in essence the same for confidence intervals as it is for tests of hypothesis.

To make this clear, assume that the student has no misunderstanding at all in setting up either of those procedures. For confidence intervals the student understands how the underlying process comes about, with the interplay between sample size, confidence and margin of error, and with the conditions needed in order to use the appropriate distribution. In a similar way, the student also understands the reasoning for setting up a statistical test, including the level of significance, the power, the sample size and the effect size, together with the necessary conditions and the reasoning "under the null hypothesis". All this is about "setting up the underlying chance mechanism". Indeed, this is the task of the statistician, and this is, after all, why students learn statistics. They should realize that deciding "with which chance process one has to work" is a scientific investigation which is completely under the control of the statistician, and under his responsibility.

Then comes the "use" of that model. A sample is taken (in the appropriate way) and "one particular realisation" of the model is computed. This is now a well-defined (fixed) interval or it is a quantity that, yes or no, falls in the rejection region. On the basis of this single result "a decision" is taken. At this point, there is nothing random anymore. The decision taken, after the result is seen, is either right or wrong. The statistician knows this and has to live with it. The fact that taking a decision based on a single outcome (of a chance process) is based on "science" and not on "mere guessing" stems from the procedure which has been set up "beforehand". "Afterwards" there is only the outcome that came to you, and the decision you have to take based on that evidence. You can hope for the best, but you never know.

REFERENCES

Agresti, A. and Franklin, C.A. (2007). *Statistics: the art and science of learning from data.* Upper Saddle River, NJ: Pearson Education, Inc.

Biehler, R. (1997a). Software for Learning and for Doing Statistics. *International Statistical Review*, **65**, 167-189.

Biehler, R. (1997b). Students' Difficulties in Practicing Computer-Supported Data Analysis: Some Hypothetical Generalizations from Results of Two Exploratory Studies. In J. Garfield & G. Burrill (Eds), *Research on the role of technology in teaching and learning statistics* (pp. 169-190). Voorburg, The Netherlands: International Statistical Institute.

Cobb, G.W. and Moore, D.S. (1997). Mathematics, Statistics, and Teaching. *Amer. Math. Monthly* **104**, 801-823.

Pratt, D. (1998). Expressions of Control in Stochastic Processes. *Proceedings of the Fifth International Conference on Teaching of Statistics. Vol 2*. Voorburg (NL). The International Statistical Institute.

Pratt, D. (2000). Making Sense of the Total of Two Dice. *Journal for Research in Mathematics Education,* **31**, 602-625.

Watkins, A.E., Scheaffer, R.L. and Cobb, G.W. (2004). *Statistics in action. Understanding a world of data.* Emeryville, CA: Key Curriculum Press.

Yates, D.S., Moore, D.S. and Starnes D.S. (2003). *The practice of statistics.* New York: W.H. Freeman and Company.

CONDITIONAL PROBABILITY PROBLEMS AND CONTEXTS. THE DIAGNOSTIC TEST CONTEXT

Marta Carles, *IES Cheste, Valencia, Spain*

M. Pedro Huerta, *Universitat de València, Spain*

ABSTRACT

In this paper we present an analysis of didactic phenomenology (Freudenthal, 1983) of the conditional probability in a particular context. The analysis takes place in a specific setting - a health setting, and also in a particular teaching environment, namely, teaching upper levels of conditional probability in a secondary school. The main aim of this work is to analyze the phenomena that are present in ternary problems of conditional probability used in the aforementioned environment. The main purpose of this work is to provide teaching professionals with some didactical elements for reflection on the use of conditional problem solving in-context for teaching this topic to secondary school students, so that we can help to improve their conditional probability literacy and skills.

INTRODUCTION

Problem solving and in particular probability and conditional probability problem solving are topics that are usually taught around the world with greater or lesser degrees of success. Some time ago Shaughnessy (1992) pointed out the difficulties of teaching probability by relating them to teaching problem-solving because teaching probability and statistics is teaching problem solving, he said.

On the other hand, in our country (BOE, 2001; DOGV, 2002) and also in other countries (NCTM, 2000), curricular standards suggest that in general mathematics, and also probability and consequently probability problem solving, should be taught in context, including mathematical context and connecting school mathematics with experimental reality of students.

Some types of conditional probability problems (for example P1 to P4 in the annex), those that Cerdán & Huerta (in press) call *ternary problems of conditional probability*[1], can be theoretically classified in different problem types (Yáñez, 2001)

---

1 In short, *Ternary Problems of Conditional Probability* are problems in which in the text of the problem at least:

1. A conditional probability is involved, either as data or as a question or both.
2. Three quantities are known.
3. All quantities, both known and unknown, are related by means of ternary relationships such as these: $P(A) + P(noA) = 1$; $P(A \cap B) + P(A \cap noB) = P(A)$ (Additive relationships); and $P(A|B) \times P(B) = P(A \cap B)$
4. Question in problem is an unknown quantity that is related to other quantities by means of more than one the relationships above.

depending on the data given in the text of the problems. Using this classification, it has been experimentally observed (Lonjedo, 2003; Lonjedo & Huerta, 2004; Huerta & Lonjedo, 2005) that through secondary school textbooks not all of these types of problems are being used in teaching. Therefore, teaching conditional probability is based on a few types of problems while ignoring others that could significantly improve the students' understanding of this subject.

In this piece of paper we will show the first steps involved in a broader research project. More specifically, we will show an analysis of didactical phenomenology of problems formulated in non-mathematical settings. This is in order to better the students' understanding and the meanings of the concepts related to conditional probability.

How to carry a context-based teaching of conditional probability out should be another research question that must be answered in the future.

BACKGROUNDS

In other pieces of work (Huerta & Lonjedo, 2003; Cerdán & Huerta, in press) the subject of this study, i.e., conditional probability problems and ternary problems of conditional probability, have already been defined. A mathematical reading of these problems allows us to classify them by means of a three-component vector (x, y, z) which represent the data in the text of the problems: x represents the number of absolute probabilities, y the number of intersection probabilities, and z the number of conditional probabilities and $x+y+z=3$. By choosing x, y and z in a suitable way we are theoretically able to identify 9 types of conditional probability problems. Only when one is competent with the algebraic register (Yáñez, 2001; Cerdán & Huerta, in press) is one capable of solving all of these different problem types. Huerta & Lonjedo (2005) showed that there are some types of problems that are not included in secondary school textbooks. Cerdán and Huerta (in press) use trinomial graphs (Fridman, 1990) in order to study ternary problems of conditional probability by means of the analysis and synthesis method.

Evans and others (2000), Girotto and González (2001), Hoffrage and others (2002), Huerta and Lonjedo (2006), Lonjedo & Huerta, (2007), Maury (1984) and Ojeda (1996) study aspects that have an influence on students' behavior in solving conditional probability problems. But, we do not know of any work in which the purpose of the research is focused on the study of contextual and phenomenological aspects of probability problems (according to Freudenthal, 1983). We think that this kind of study is necessary in order to find out if contextual factors also have an influence on problem solving processes or which contexts we should use in order to teach conditional probability to enhance students' understanding of the subject.

OBJECTIVES AND METHOD

The main objective of this paper is to show a phenomenological analysis of conditional probability problems as described above. More specifically, we will analyze the problems related to teaching–learning environments where conditional probability is being taught. These environments do not include secondary schools because in order to answer the question we put forward in the introduction, a positive reply will involve, among other reasons, the fact that secondary school students today will find this type of problems in their futures both in further education and working environments.

All work we are reporting in the paper is made thinking in secondary school. If we think in preparing students in conditional probability in secondary school, at least, two questions can arise: why? and how? The answer to the first one takes in account students' future, both as a student in a university and as a citizen. So, it is necessary to explore contexts and phenomena in which conditional probability is involved. If we do think in this way, we will determine not only what kind of competences secondary schools students must have with conditional probability, but also what type of problems they have to solve and in which contexts problems have to be stated. On the other hand, the answer to the second question can be found in a phenomenological and realistic approach to teaching conditional probability through problem solving.

Due to the fact that we investigate problems in teaching-learning environments, our information comes from several sources. One of them is from textbooks in Colleges at Universitat Politècnica de València and Universitat de València and another one is from the Internet, introducing a word chain in a searcher as follows: Probability, Conditional Probability, Conditioned Probability, Bayes' Theorem, and so on. In both cases, the main item in the search was conditional probability problems that have been used in teaching during the 2005-2006 school year. Furthermore, whenever possible, we interviewed professors from the aforementioned Universities or get into contact with specialist people in some topics related to ours, or sent e-mails to the other teachers and professors at the other information sources. Occasionally, through the use of the Internet we found papers, research projects or Doctoral Dissertations that used conditional probability in different contexts other than those related to the teaching of Mathematics.

By analyzing the aforementioned documents we were able to classify them according to the following criteria: Context (in which the problem is formulated), Phenomena referring to events (that is to say, organized by means of events), Phenomena referring to probabilities (that is to say, organized by means of probabilities), Specific terminology, Classification (attending to the structure of the data in the text of the problem and the presentation format of the data) and Specific teaching environment or reference. We then went on to define the aforementioned criteria as follows:

Context: A particular situation in which problems are put forward. In a context, a particular concept such as conditional probability has a specific meaning or is used with a specific sense. For example, a Diagnostic Test is a context.

Phenomena (referring to events): In a particular context, those statements that can be recognized as having an uncertain possible outcome are phenomena. These statements can be organized by means of references sets (Freudenthal, 1983, p.41) - events in a probabilistic and mathematical sense and operations between events. For example, “being ill”, “being ill and having a negative diagnostic”. Neither of these phenomena will be recognized as a “conditional event” even though it is possible to talk about them as if they were. For example, “knowing that he /she has a negative diagnostic, being ill”.

Phenomena (referring to probabilities): In a particular context, a part from quantities, we refer to signs, words, sets of words and statements that express a measurement or the need for a measurement regarding the uncertainty of a phenomenon. For example, in the phrase *risk of “false alarms”, false alarms* could be recognized as an event but, mathematically, it is not an event. However, *risk* indicates that something is probably wrong and has possible undesirable consequences. There a measurement of risk is related to it. In the Diagnostic Test context, *risk of “false alarms”* is sometimes called *FPC or False Positive Coefficient.* Prevalence of a disease is another example of phenomena we are referring to. The encyclopedic meaning of the word “prevalence” is not related to probabilities. But in some cases, when this word is used tied to a particular context, as we are considering it in this work, it acquires a particular meaning in a probabilistic sense. So, the phenomenon of *prevalence of a disease* can be organized by means of a probability and expressed in the way that it is usual: frequencies, percentages or number in [0, 1].

Dictionary of specific terms: In a particular context, and concerning a particular problem, signs, words, sets of words, statements and their meanings within the given context. Sometimes, most of these meanings are already familiar terms.

Specific teaching environment or reference: College, University, or whatever other teaching Centres from which the problem has originated.

RESULTS AND DISCUSION

Taken into account the items for the analysis mentioned above, for the problems listed in the annex we have drawn up this table:

Problem	Context	Phenomena (Events)	Phenomena (Conditional probability)	Specific Terms	Problem Classification (Known data and question)	Data Format.	Specific teaching environment or reference
P1	Diagnostic Test in health setting	-Tubercular / Non-tubercular. -A tubercular / Non-tubercular person can give positive / negative in test. -A positive person tested can suffer / not suffer from tuberculosis.	Reliability of test for the disease diagnostic.	-To give positive / negative in test. -To be / not to be ill. -PPV or Positive Predictive Value	(1,0,2) P(D|+)	Rate. Percentages	Hospital Universitario Ramón y Cajal de Madrid. http://www.hrc.es/bioest/Probabilidad_18.html
P2	Diagnostic Test in health setting	The same, changing tubercular person by Prevalence of diabetes.	Reliability of test for the disease diagnostic.	-FPC or False Positive Coeficient -FNC or False Negative Coeficient -PPV or Positive Predictive Value -NPV or Negative Predictive Value	(1,0,2) p(D|+) p(no D|-)	Percentages	Hospital Universitario Ramón y Cajal de Madrid. http://www.hrc.es/bioest/Probabilidad_18.html
P3	Diagnostic Test in health setting	The same, uterine cancer	Reliability of test for the disease diagnostic.	-False Positive Coefficient (FPC) -False Negative Coefficient (FNC) -Pre-test probability -NPV or Negative Predictive Value	(1,0,2) P(no D|-)	Probability	Hospital Universitario Ramón y Cajal de Madrid. http://www.hrc.es/bioest/Probabilidad_18.html
P4	Diagnostic Test in health setting	Prevalence of the tuberculosis	Reliability of test for the disease diagnostic.	Sensitivity Specificity False positive To be ill	(0,0,3) P(D)	Probability	Faculty of Mathematics, Universitat de València. Probability and stochastic processes.

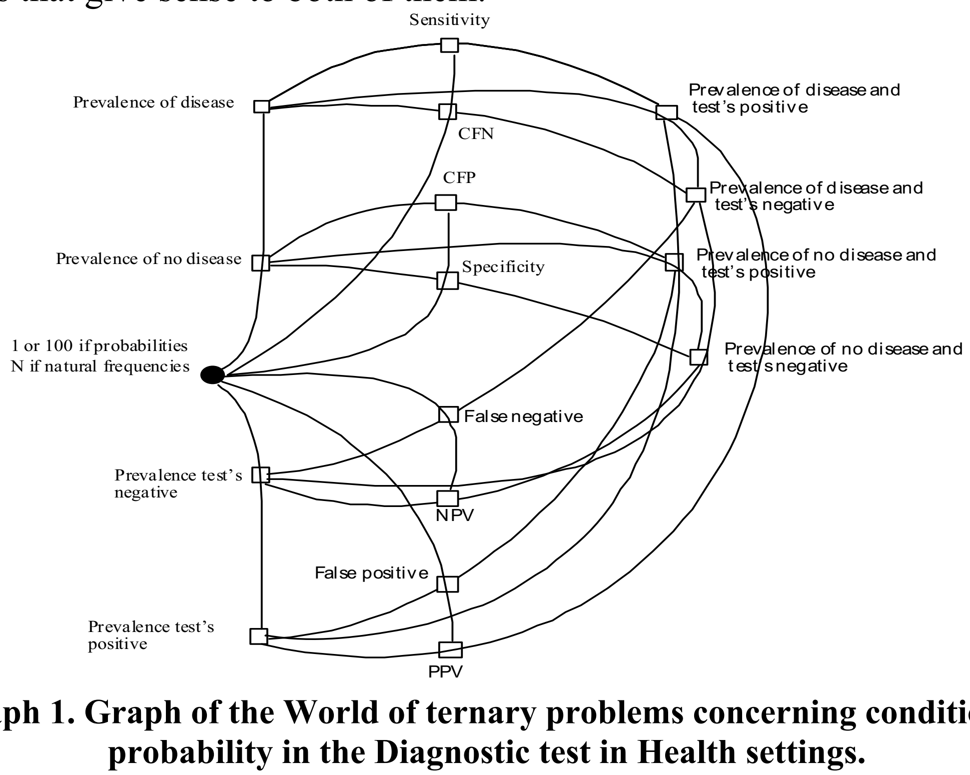
Table 1:Some results of the phenomenological analysis of problems 1 to 4.

In the Diagnostic Test context we can also highlight several situations in which conditional probabilities are used, depending on the purpose the test is to be used for: diagnosis of diseases, diagnosis of faulty articles or products in manufacturing, diagnosis in legal Biology, and so on. In particular, problems 1 to 4 are stated in a Diagnostic Test context in a health setting.

Let M_i be, i= 1, 2; the two possible values for the mistakes produced in disease diagnostic, and S_i, i=1, 2 the two possible values for the successes in the disease diagnostic (or predictive values). We have the following expressions to calculate the aforementioned values: M_1= probability of one kind of mistake induced by a test (the FPC), M_2= probability of the other kind of mistake induced by a test (the FNC). S_1= probability of one kind of success produced by a test (the sensitivity), S_2= probability of the other kind of success produced by test (the specificity).

And also the following relationships: M_1= FPC x (1-prevalence), M_2=FNC x (prevalence), S_1= sensitivity x prevalence / (sensitivity x prevalence + FPC x (1-prevalence)), S_2= specificity x (1-prevalence)/ (specificity x (1 – prevalence) + NFC x prevalence)

We have been able to make a reading of the mistakes and success using probabilities. Therefore, in this sense, we have 8 conditional probabilities and 4 absolute or marginal probabilities, complementary in pairs. Some of these complementary relationships are important in this context: FPC + specificity = 1; FNC + sensitivity = 1 The next graph represents the world of ternary problems of conditional probability (Cerdán & Huerta, in press) in the context we are considering in this paper. The graph shows all relationships between events and quantities that give sense to both of them.



Graph 1. Graph of the World of ternary problems concerning conditional probability in the Diagnostic test in Health settings.

Consequently, in terms of ternary problems in conditional probability, and in this context, as there are no intersection probabilities, we can reasonably state problems organized by vectors (x, 0, z) with x+z= 3, *x* being the number of absolute probabilities and *z* the number of conditional probabilities in the text of the problem, and choosing them in an appropriate way.These types of problems are (1, 0, 2), (2, 0, 1) and (0, 0, 3). Problems 1 to 3 in the annex belong to the first type and they ask about mistakes or successes in disease diagnostics. Apart from the third type, all of these problems have arithmetical resolutions (Yañez, 2001; Lonjedo & Huerta 2004; Huerta & Lonjedo, 2005). Type (0, 0, 3), however, organizes problems that have an algebraical solution using explicitly FPC (FNC), Sensitivity (Specificity) and a Mistake (Success) as data in the text of the problems and they must ask about an absolute probability (prevalence of a disease or positive test result).

Finally, let us suppose a particular situation organized by (2, 0, 1). Let x% be the sensitivity of test, *a* the prevalence of the disease and *b* the prevalence of the test's positive results. We suppose S_1 is the question. The following proportion is the answer: $S_1/x=a/b$, $S_1=xa/b$, expressed necessarily by a percentage.

It is possible to make a similar analysis of the requirements for data and the relationships between data for the other types of problems. In general, but depending on the question in the problem, it could be stated that when the number of conditional probabilities in the text of the problem increases, the number of relationships also increases, thereby making the problems more complex in structure.

CONCLUSIONS

The problems that we are considering in this study may be considered as problems of application for conditional probability, that is to say, problems that are solved after formal teaching of the concept of conditional probability have been carried out. But, teaching of this topic based on an exploration of the phenomena, that involves its uses in different contexts, is precisely the opposite point of view from formal teaching. Hence, as Freudenthal (1983) suggests, teaching the topic of conditional probability in secondary school might begin with solving problems that let students make an exploration of the phenomena involved with the topic and which could then be followed by teaching the formal topic of the conditional probability as a means of organizing those phenomena. The problems we have analyzed in this paper could be an example of this. These problems should be dealt with in secondary school education including contexts as far as possible. And only College level students should be taught about the formal concepts of conditional probability as a means of modeling the concepts within a context.

One of the most commonly used contexts has been analyzed in this work. We termed it as the Diagnostic Test. It can be recognized in several different settings:

in textbooks, in research on students' behavior in solving conditional probability problems and so on. Generally, data and the relationships between data are not previously analyzed in relation to the context. But, if, when we think about teaching conditional probability we previously analyze problems as we suggest in this work, we can determine what type of problems can be reasonably proposed to our students at every level of education that the subject matter is taught and in which context those problems must be stated in order to improve students' understanding of conditional probability.

REFERENCES

B.O.E., (2001). Real Decreto 3473/2000, de 29 de diciembre, por el que se modifica el Real Decreto 1007/1991, de 14 de junio, por el que se establecen las enseñanzas mínimas correspondientes a la Educación Secundaria Obligatoria. *Boletín Oficial del Estado*, nº 14. Madrid: Ministerio de Educación y Ciencia.

Cerdán, F., Huerta, M.P., (in press). *Problemas ternarios de probabilidad condicional y grafos trinomiales.* Educación Matemática.

D.O.G.V., (2002). Decreto 39/2002, de 5 de marzo, del Gobierno Valenciano, por el cual se modifica el Decreto 47/1992 de 30 de marzo, del Gobierno Valenciano, por el cual se establece el currículo de la Educación Secundaria Obligatoria en la Comunidad Valenciana. *Diari Oficial de la Generalitat Valenciana*, nº 4206. Comunidad Valenciana: Conselleria de Cultura i Educació.

Evans, J., Handley, S.J., Perham, N., Over, D. E. & Thompson, V. A., (2000). Ferquency versus probability formats in statistical word problems. *Cognition 77,* pp. 197-213.

Fridman, L.M., (1990). *Los grafos Trinomiales como Metalenguaje de los Problemas, en matemáticas. Revista del Departamento de Matemáticas de la Universidad de la Sonora*, 17-18, pp.51-59.

Freudenthal, H., (1983). *Didactical Phenomenology of Mathematical Structures*. Dordrecht: Reidel.

Girotto, V. & González, M., (2001). Solving probabilistic and statistical problems: a matter of information structure and question form. *Cognition 78*, pp. 247-276.

Hoffrage, U., Gigerenzer, G., Graus, S. & Martignon, L., (2002). Representation facilities reasoning: what natural frequencies are and what they are not, *Cognition 84*, pp. 343-352.

Huerta, M. P., Lonjedo, M. A., (2003). La resolución de problemas de probabilidad condicional. In Castro, Flores at alli… (eds), 2003, *Investigación en Educación Matemática. Séptimo Simposio de la Sociedad Española de Investigación en Educación Matemática*. Granada.

Huerta, M.P., Lonjedo, M. A., (2005). Los problemas de probabilidad condicional en la Enseñanza Secundaria. *XI Jornadas sobre el aprendizaje y la Enseñanza de las Matemáticas (XI JAEM)*, edición en CD-ROM. Colección: Encuentros Educativos. Canarias: Consejería de Educación, Cultura y Deportes del Gobierno de Canarias.

Huerta, M.P., Lonjedo, M. A. (2006) The Nature of the quantities in a conditional probability problem. Its influence in the problem solving behaviour. In M. Bosch (ed) (2006), *European Research in Mathematics Education IV. Proceedings of the Four Congress of the European Society for Research in Mathematics Education*, pp. 528-538. Sant Feliu de Guíxols, Spain.

Lonjedo, M. A., (2003). *La resolución de problemas de probabilidad condicional. Un estudio exploratorio con estudiantes de bachiller*. Departament de Didàctica de la Matemàtica. Universitat de València (Memoria de tercer ciclo no publicada)

Lonjedo, M. A., Huerta, M. P., (2004). Una clasificación de los problemas escolares de probabilidad condicional. Su uso para la investigación y el análisis de textos. In Castro, E., & De la Torre, E. (eds.), 2004, *Investigación en Educación Matemática. Octavo Simposio de la Sociedad Española de Investigación en Educación Matemática*, pp 229-238, Universidade da Coruña. A Coruña.

Lonjedo, M.A., Huerta, M.P., (2007). *The same problem in three presentation formats: Different percentages of success and thinking processes*. Paper presented in CERME 5. Larnaca, Cyprus, February 22-26.

Maury, S., (1984). La quantification des probabilités: analyse des arguments utilises par les élèves de classe de seconde, *Recherches en Didactique des Mathématiques,* vol. 5, nº2, pp. 187-214.

Ojeda, A.M., (1996). *Representaciones, contextos y la idea de probabilidad condicional*, en Investigaciones en Matemática Educativa, pp. 291-310. (Editor F. Hitt). Grupo Editorial Iberoamérica. México.

NCTM, (2000). *Principles and Standard for school mathematics*. Reston, VA; N.C.T.M. On line: http://standards.nctm.org/

Shaughnessy, J.M., (1992). Research in Probability and Statistics: Reflections and Directions, en D. Grouws (ed.) (1992), *Handbook of Research on Mathematics Teaching and Learning,* pp. 465-494. (MacMillan Publishing Company: New York).

Yáñez, G., (2001). El álgebra, las tablas y los árboles en problemas de probabilidad condicional, en Gómez, P., y Rico, L. (eds.), *Iniciación a la investigación en didáctica de la matemática. Homenaje al profesor Mauricio Castro,* pp.355-371. (Editorial Universidad de Granada: Granada).

ANNEX

P1. It is known that in a certain city one out of every 100 citizens is a tubercular person. A test was administered to a citizen. When a person is tubercular the test gives a positive result in 97% of cases. When he/she is not tubercular, only 0.01% of the cases give positive results. If the test is positive for these people, what is the probability that he/she is tubercular?

P2. A diagnostic test for diabetes has an FPC of 4% and an FNC of 5%. If the prevalence of diabetes in a town is 7%, what is the probability that a person is diabetic if his/her test was positive? What is the probability that a person is not diabetic if his/her test was negative?

P3. A diagnostic test for uterine cancer has a false positive coefficient of 0.05 and false negative of 0.01. A woman with a pre-test probability of 0.15 of having the disease has a negative result in her test. Calculate the probability that she is not ill?

P4. The tuberculin test can test whether a person is infected by tuberculosis or not. The sensitivity and specificity of the test is very high, 0.97 and 0.98 respectively. If in a certain town there is a very high proportion of false positives, exactly 0.9, calculate the prevalence of the disease.

A MICROWORLD TO IMPLANT A GERM OF PROBABILITY

Cerulli Michele, Chioccariello Augusto, Lemut Enrica

Istituto Tecnologie Didattiche - C.N.R. di Genova

This paper reports on the third part of a long term experiment concerning the introduction of 12-13 years old pupils to the concept of randomness and probability. In the first parts, the pupils analysed the concept of randomness, in daily life and with LEGO-RCX robots. The third part is based on the Random Garden Game (within the programming environment ToonTalk). The pupils are guided towards the concepts of equivalence of sample spaces and of classical probability, through temporary and personal meanings, corresponding to an evolution by means of class discussions.

INTRODUCTION

This paper accounts of some findings of a long term experiment [1] based on technological tools to introduce pupils to the concepts of randomness and probability. The experiment is composed of three phases: the first two concern pupil's introduction to the concept of randomness (Cerulli et al., 2006). The third phase focused on the concept of probability, with particular attention to: **a)** Pupils' development of the concept of sample space and of "equivalence" with respect to the events involved; **b)** Pupils' construction of theories for comparing sample spaces; **c)** Pupils' construction of the concepts of frequency, relative frequency and probability.

The focus of this paper is mainly on how we used the computer programming environment ToonTalk (Kahn, 2004) to approach issue **a)**.

THEORETICAL ASSUMPTIONS

Research literature shows difficulties related to pupils' introduction to probability (Fischbein, 1975; Pratt, 1998; Wilensky, 1993; and Truran, 2001), witnessing the failures of standard approaches. One of these difficulties concerns pupils' capability of interpreting different random phenomena according to a unifying perspective (Pratt et al., 2002; Nisbett, 1983 via Pratt). In order to avoid this difficulty, our experiment is based on the idea of presenting pupils with different random phenomena that teachers and pupils can recognize and consider as belonging to a sector, unitary and homogeneous, of the human culture, that we can identify as *experience field of aleatory phenomena* (Boero et. Al., 1995). A unifying perspective is partly achieved in the first phases of the experiment, where phenomena such as the tossing of a coin, and one-dimensional random walk, were represented by means of unique robot, the Drunk Bot (Cerulli et al., 2005). However, according to Pratt (1998) the construction of meaning within this subject lies in connecting its formal and informal views. A formal view according to Noss and Hoyles (1995) can be achieved thanks to the introduction of suitable computer microworlds in the school practice. The activity we are presenting is based on a specific microworld, the Random Garden that we built ad hoc in ToonTalk. The key idea is that the microworld can be used as a unifying model for representing and manipulating random phenomena. This model is unifying in the sense that each random phenomenon is represented as a sample space with a random extraction process, and different phenomena's representations are structurally the same. Moreover, the analysis and manipulations of the phenomena, within the microworld, are also qualitatively identical.

According to socio-constructivism, our experiment assumes that learning can be the result of active participation in both practical and social activities. However, such kinds of activities do not guarantee that the meanings constructed by the pupils are coherent with mathematics or with the teacher's educational goals. Such coherence can be achieved by means of mathematical class discussions orchestrated by the teacher (Mariotti, 2002; Bartolini Bussi, 1996).

THE RANDOM GARDEN TOOLS

The Random Garden is a microworld, for representing random extraction processes. The tool consists of a sample space (the *Garden*) a *Bird* and a *Nest*. [2] When the user gives a number to the *Bird*, a corresponding number of objects is extracted (with repetitions) from the *Garden* and deposited in the *Nest* (Figure 1).

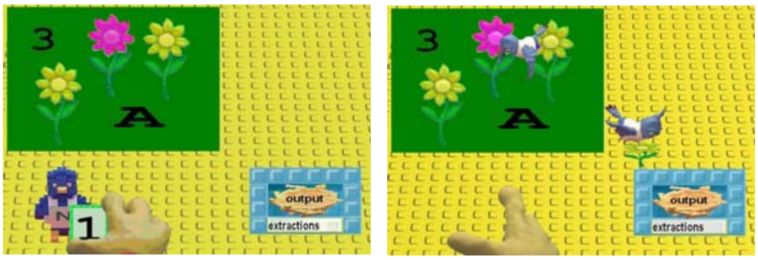


Figure 1. A number is given to the N-bird to request a random extraction of *N* objects. A new bird comes out and drops extracted objects in the output nest.

Figure 2. Eight extractions are collected in a box containing a nest (left); only the first element is clearly visible. The nest can be converted into a box with eight holes showing the ordered sequence of extracted objects (right).

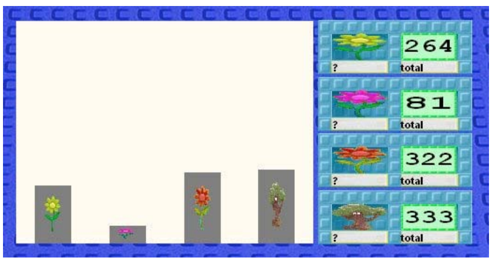


Figure 3. The *Bar Graph* (left) and the *Counters* (right), show respectively the proportions and the exact numbers of elements extracted for each kind of object.

The user can modify the garden by adding or removing objects which can be numbers, text, or images of any kind. It implies that this simple device can be used as a means for representing any kind of random phenomena. The elements extracted from the Random Garden are collected in a box containing a nest (Figures 1 and 2).

In order to visualize the whole sequence of extractions, it is possible to convert the nest into a new box with as many holes as the number of extracted objects (Figure 2). Hence, a rough qualitative view of the sequence and of its properties can be seen at a glance. If the number of extractions is large, and/or if one needs a more detailed qualitative/quantitative analysis of the data, other tools are required and provided: *Bar Graph* and *Counters* (Figure 3). These are dynamic tools, in fact numbers and bars change while the extraction is in progress. In particular, the bars oscillate at the beginning of the extraction process, and stabilize after a large number of extractions.

THE GUESS MY GARDEN GAME

If one is given a nest, or a box of extractions (see Figure 2), it is possible to address the question of what a possible composition of the Random Garden that generated the given sequence is. This key question is at the core of the Guess my Garden game [3] which is conducted as follows. One team of pupils, namely a small group, creates a *Random Garden* (using at most 12 objects), thus defining a sample space; then produces a set of boxes containing increasing numbers of extractions, for example, 2 boxes with 100 extractions and 2 boxes with 1000 extractions, and so on.

All the boxes, containing the data generated by the pupils' team, are included in a ToonTalk *notebook* named like team, and the *notebook* is published on the web, as a challenge for other players. Another team can then download the notebook and analyse the data it contains in order to try to guess the makeup of the Random Garden produced by the challenging team. The team can either simply observe the sequences of extractions, or study them using the *Bar Graph* and *Counters* tools. Once they make a conjecture concerning the garden to be guessed, they can produce a new corresponding garden and use it to produce a number of extractions that may be compared to those provided by the challenging team. Once the team is satisfied with the conjectured garden, they can publish it on the web and wait for their counterparts to validate or invalidate their answer. Finally, the challenging team checks the published answer and posts a comment to inform the other team whether they have guessed their garden correctly or not. If the garden has not been guessed, then the exchange between the pupils can continue until an agreement is reached.

THE EPISODE

The first two phases of experiment involved reflective and practical activities aiming at exploration and consolidation of some key issues related to randomness (eg. *unpredictability, fairness, indeterminism, random walks, etc.*). In the first phase pupils collected, proposed, and analysed sentences, talks, and episodes related to randomness. Pupils had to write individual reports and to discuss some of the emerging items with the rest of the class. Each considered item was discussed in terms of key questions such as "is it random or not?", or "is it predictable or not". The results of the discussions were reported in a shared class document, called *Encicloaedia of Randomness*. The second phase was based on the interaction with some LEGO robots that incorporated different aspects of the concept of randomness (see Cerulli & al. 2006, for more details). Each robot was discussed and classified as random or not random: the same questions used in phase one functioned as pivots for the teacher's orchestration of the discussions. After these phases, the pupils were very familiar with the concept of randomness and were about to be introduced to probability starting from the Guess my Garden game.

A class of 21 pupils from Milan starts the game by making public a set of challenges, and receiving answers from Swedish and Portuguese opponents. The first protagonists of the episode are the members of Jeka's team: Jeka (Jk), Jè (Je) and Rossana (R). They have to build a *garden* to be used to publish a difficult challenge as required by the task (Italian dialogues translated by the authors):

Jk:	we could do…the same number…of flowers and trees
Jk	and Rossana: three times this one, three times this one, three times this one and three times this one (*pointing to the objects in the random garden*)

The girls build their garden and ask the software to produce 100 extractions from the garden, and then comment on the results:

Jk:	yellow flower 25 extracted times…
R:	…but they are all the same?! (*looks surprised*)…more or less…25, 24, 25 and 26 …ah, yes, of course, we put (*in the garden*) all the same numbers (*of flowers and tree*)…(*she looks around to stress that she is stating something obvious and her pals nod*).

The girls take note of the obtained result and a researcher (M) intervenes to investigate what strategy the pupils are using to build a difficult challenge:

R:	we multiplied each object of the garden by 3 (*pointing to the monitor*)…we tripled
M:	Why do you think this is difficult to be guessed? (*reads one of the written questions pupils are supposed to answer in order to accomplish their task*)
R:	no, it is not difficult, we just tried…
M:	but it is not easy to guess this ...*(he is promptly interrupted by Jeka)*
Jk:	exactly! Because…one may think of two (*objects*) maybe…
R:	…yes…. (*thoughtful*)
Jk:	I would think of two (objects of each kind in the garden)

The original idea (to triplicate the numbers of flowers) of Jeka begins to be clearer and becomes more explicit when the teacher (A) asks them for an explanation:

A:	why do you think this garden is difficult…?
Jk:	maybe because with the resulting numbers *(after the extraction)* one may…one may get confused

A: why? What answer could you get?
Jk: maybe two
M: you mean two…
Jk: I mean, if one sees 20
A: 2, 2, 2, 2? *(meaning a garden with 2 objects of each kind; Jk, Jè and R nod)*

After some reflections and discussions, the girls decide to publish the garden they had produced, made of 3 objects of each kind (Figure 4). They believe that their opponents may think that the garden is made of 1 object (or 2 objects) of each kind.

Figure 4. On the left it is possible to see the garden defined by Jeka's team, while on the right it is possible to see the attempt of guess sent by a Swedish team.

The following week the class goes back in the computer laboratory and each team finds an answer from a Swedish team. Jeka's team finds the answer given by Amelie's team (Figure 4), who conjectured that the *garden* contained 1 object for each kind, instead of 3, as expected by Jeka's team. The answer is considered wrong by Jeka's team, as they write in the message they send to the Swedish opponents.

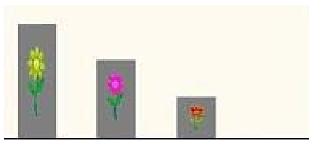


Figure 5. The results of 1200 extractions from M's challenge.

After responding to their Swedish pals, the Italian pupils are required to discuss how to respond to M's challenge. Such challenge is analysed by two teams who obtain the graph of Figure 5 using the *Bar Graph* tool.

Figure 6. On the right: Lollo's team's garden; on the left Jeka's team's garden.

The two teams that answer to M's challenge give different guesses: a) 2 red flowers, 4 pink flowers and 6 yellow flowers (Jeka's team); b) 1 red flower, 2 pink flowers and 3 yellow flowers (Lollo's team).

Which of these two answers is correct? The teacher poses this question a couple of weeks later within a class discussion where the pupils couldn't access the computers and the Random Garden tools. In order to support the class discussion, the teacher provides a set of cards representing the gardens proposed by the two teams, as shown in Figure 6.

The teacher recalls the two teams' different answers and asks the class which of them is right or wrong, or if they are both wrong or right. A brief discussion follows where pupils agree that the chances to get a red flower from garden A or from garden B are the same, as expressed by C1:

C1:	they are the same; in fact they are all doubled (*meaning that the number of flowers in garden B is twice the number of flowers in garden A*)

The gardens are now regarded as objects and their constitutive properties are compared, and garden B is now regarded as a sort of "double" of garden A. This idea reminds Jeka of the challenge they proposed to their Swedish pals.

Now the terrain seems to be ready for planting the seed of equivalence and the teacher takes this occasion to introduce explicitly the word "equivalent":

A:	[…] thus these two gardens, in theory, are equivalent?
C1:	yes, they are equiv…
A:	[or] Are they equal?
C2:	they are equal
A:	[or] Are they identical?
C1:	they are equal
Jk:	they are equivalent
C1:	they are equiv…

A noise of chat among pupils follows which ends with C1 stating:

C1:	equivalent!
Jk:	they are equivalent! (*Someone in the background says "equal"*)
C1:	equiva…equivalent...*(seems to be doubting)*
A:	will you explain me what you mean by equivalent?
Jè:	not equal because there are not the same elements in the two gardens
A:	Thus M surely had one or the other *(meaning that if M had one of the proposed gardens, he couldn't have both of them but only one)*
Jè:	equivalent because they have the same values….in practice…we can say so!

After a while, the position expressed by Jè (when she states that the two gardens are not equal because they have different elements) opens the space for discussing new criteria (different from pure "equality") for comparing gardens. Jè introduces her criterion of equivalence, based on the "values" [4] of the gardens. At this point it is not clear what is meant by the word "values", but from the context we deduce that Jè refers to some kind of result produced by the gardens that it could be the graphs, the numbers in the *Counters*, or the sets of extractions. In the meantime, other pupils, like Bo, propose their interpretation or alternative "definition" of equivalence:

Bo: the percentages are the same…I believe it is because the percentage is more or less the same.

A: the same of what?

Bo: of…of twelve…of all the flowers…for instance

A: give me an example

Bo: in the first garden, the garden A, the percentage is 6…in the first garden the percentage is 6, six is …it is …

A: the percentage? The total?

Bo: yes, the total, the 6 is like the 100, and the red is 1, thus it is ….the red is 1 over 6…oh god!

A: 1 over 6

Bo: 1 over 6 and the second (garden) is 2 over 12 which is like 1 over 6

A: uhm, thus you say "they are equivalent as the two fractions" that you said?

Bo: yes

A: right?

A: but they are not the same, so we can say that these two answers are equivalent […] but without knowing Michele's garden we cannot give a definite verdict, ok?

Bo: because even if we could make the extraction….surely the numbers would not be the same, they would be almost the same!

A: almost the same, and the colums [of the graphs]?

Bo: equ….with the same heights I think

A: in the two gardens?

Bo: in the two gardens the heights of the columns would be the same but the values different

Bo's idea of equivalence, different from Jè's, is based on an analysis of the constitutive elements of the garden, and he tries to formalize it by associating fractions to the garden. The fractions are only by chance coherent with the classical definition of probability which at this stage is not known to these pupils [5]. However, Bo also agrees that the two gardens would produce the same graph.

The class needs to establish criteria to validate responses to challenges. This leads them to introducing some idea of equivalence. Such idea is still fuzzy, but nevertheless it turns out to be useful in the following excerpt where the class discusses the case of Jeka's team and their Swedish opponents. According to Jeka, the strategy used by M is the same as the one used by her team. They both produced a garden whose results were "the same" as the results of other gardens, so that their opponents' chances to guess are low. This issue relates to the idea of equivalence of gardens, so the teacher approaches it just after the discussion of M's challenge, and puts it in terms of validation of the Swedish answer:

A: they [Jeka's team] answered to the Swedish pupils "you did not guess", are we leaving this answer or are we going to write to them again?

C: I think it is better to re-write it and state that they did well but the values…

Jk:	they did wrong
C:	they did wrong but with respect to the values…they did wrong but they were right because the original value was 3, 3, 3 and 3, the one they had to guess, and they put 1, 1, 1, 1 but it is not their fault because…
Jk:	they did wrong
C:	they did right, it is only that the bars were the same, so they could put any number
Jk:	I think we should re-write it stating that they did wrong, but that the two gardens are equivalent like the other one (referring to M's challenge)

According to C, the Swedish team's answer should be considered correct because their garden produces the same graph as Jeka's team garden. This idea is re-formulated by Jeka in terms of equivalence between the two gardens, thus the word "equivalent" (ita. "equivalente") begins to be used by the pupils as a tool for validating responses to the challenges. However, the meaning associated to the word "equivalent" needs some more clarification:

A:	what is the phrase you would write to explain them…?
C:	I would write, you did wrong…should I say also the solution?
A:	suppose that what you say is going to be sent to the Swedish pupils
C:	you did…you did wrong…you didn't guess our garden but you found another one that has the same value of the one we did, apart from…

The words "values" (ita.: "valori") and "equivalent" appears as strictly tied, and their relationship is definitely cleared by C in the following excerpt:

A:	what do you mean by "value"?
C:	I mean that the graphs of the extractions are equal to the graphs of our garden
A:	perfect
C:	only, our garden had different quantities of objects
A:	so, if they give the same extractions they are equivalent
C:	there are 3 equivalent gardens…there are 2 equivalent gardens…there is our garden, 3 red flowers, 3 yellows, 3 pinks and 3 trees, your garden, one for each object, and a third one containing to elements for each object .

Finally, the meanings of the words "values" and "equivalent" are cleared and the class agrees on the following criterion for validating answers to responses: the answer is correct if the garden proposed by the responder is equal to the original garden; the answer is almost correct if the garden of the responder is equivalent to the original garden, in the sense that they produce the same graph.

What we find interesting in this story is that the "germ" of the idea of equivalence appeared in the form of the strategy employed by Jeka's team for winning the game. It then reappears in the discussion of M's challenge which was designed on purpose to exploit the ambiguities related to the equivalence of gardens; in this case the idea of equivalence appears in the form of a criterion for deciding which of the two proposed answers is correct. Each of these steps corresponded to an evolution of pupils' idea of equivalence of gardens by means of reflections and class discussions that are clearly motivated and driven by the needs of the game. The two needs that drive such evolution are basically: the need for finding a principle to validate answers; the need to produce "difficult" challenges.

CONCLUSIONS

In this paper we show how the third phase of the long experiment contributed to pupils' development of an idea of equivalence of sample spaces, which we assume to be basic for developing a definition of probability. We observe also that in this kind of activity pupils implicitly cope with matters related to the law of large numbers.

The experiment continued by involving the pupils in tasks of comparisons of sample spaces, questioning on how easy (probable) it was to pick a given flower from a given Random Garden. Starting from these tasks, new class discussions were set up, which led pupils to define their own operative strategies for comparing sample spaces, in terms of choosing the sample space which would more likely generate a specific event. All of these strategies consisted in some kind of computation associating a "result" to each sample space: the comparison of such results would help choosing the "best" sample space. One of the strategies proposed by pupils was consistent with the classic definition of probability. This made room for introducing the classic definition as a means for studying and comparing random phenomena.

The strategies proposed by pupils are, as a matter of fact, not dependent on the context, in the sense that the flowers of the random garden can be substituted with other objects/symbols. Thus the Random Garden becomes a unifying model to represent random phenomena, as suggested by the final activities of the project, where pupils used the Random Garden tools to reproduce on the screen the behaviours of the LEGO robots they experienced in the first phases of the long experiment.

NOTES

1. We acknowledge the support European Union. Grant IST-2001-32200, for the project "WebLabs: new representational infrastructures for e-learning" (see http://www.weblabs.eu.com/).

2. See a tutorial: http://www.weblabs.org.uk/wlplone/Members/augusto/my_reports/Report.2004-06-21.4151

3. Rules and data: http://www.weblabs.org.uk/wlplone/Members/augusto/my_reports/Report.2005-01-04.5407

4. Consider that the Italian "valori" (used by Jeka and here translated with "values") can have several meanings, among which it can represent either the values of the parameters of the input of a process or the results of the same process.

5. We should mention that the class had previously discussed on what strategies can be employed to decide, given two gardens, which of them is more likely to extract a certain kind of flower. The strategies proposed by pupils include "measure" of "chances to extract" of various kinds among which we find the fractions employed in this case by Bo.

REFERENCES

Bartolini Bussi M. G. (1996). Mathematical Discussion and Perspective Drawing in Primary School. *Educational Studies in Mathematics*, 31 (1-2), 11-41.

Boero, P., Dapueto, C., Ferrari, P., Ferrero, E., Garuti, R., Lemut, E., Parenti, L. & Scali, E. (1995). Aspects of the Mathematics-Culture relationship in mathematics Teaching-Learning in compulsory school. *Proceedings of PME XIX*, Recife, Vol. 1, pp. 151-166.

Cerulli, M., Chioccariello, E. & Lemut, E. (2007) Randomness and LEGO robots. (In Bosh, M. (Eds), *Proceedings of CERME 4*. IQS Fundemi Business Institute, Sant Feliu de Guixols, Spain, ISBN: 84-611-3282-3)

Fischbein, E. (1975). *The Intuitive Sources of Probabilistic Thinking in Children*. (Reidel, Dordrecht)

Kahn, K. (2004). ToonTalk-Steps Towards Ideal Computer-Based Learning Environments. (In Tokoro, M. & Steels, L. (Eds.), *A Learning Zone of One's Own: Sharing Representations and Flow in Collaborative Learning Environments*. Ios Pr Inc.)

Mariotti, M. A. (2002). Influences of technologies advances in students' math learning. (In English, L. D. (Eds.), *Handbook of International Research in Mathematics Education* (chapter 29, pp. 757-786). Lawrence Erlbaum Associates publishers, Mahwah, New Jersey.)

Nisbett, R., Krantz, D., Jepson, C. & Kunda, Z. (1983). The Use of Statistical Heuristics in Everyday Inductive Reasoning. *Psychological Review*, 90(4), 339-363.

Noss, R. & Hoyles, C. (1996) *Windows on mathematical meanings learning cultures and computers*. (Kluwer Academic Publishers, Dordrecht/Boston/London.)

Pratt, D. (1998). The Construction of Meanings IN and FOR a Stochastic Domain of Abstraction. Dissertation University of London, Institute of Education.

Pratt, D. & Noss, R. (2002). The Microevolution of Mathematical Knowledge: The Case of Randomness. *The Journal of The Learning Sciences*, 11(4), 453-488.

Truran, J. M. (2001). The teaching and Learning of Probability, with Special Reference to South Australian Schools from 1959-1994. Dissertation, University of Adelaide.

Wilensky, U. (1993). Connected Mathematics-Building Concrete Relationships with Mathematical Knowledge. Dissertation, MIT.

THE IMPACT OF A TYPICAL CLASSROOM PRACTICE ON STUDENTS' STATISTICAL KNOWLEDGE

Andreas Eichler

Universität Münster

This report focuses on a research project that combines three aspects of a curriculum concerning teachers' planning, teachers' classroom practice, and their students' statistical knowledge. Firstly, the theoretical framework and methodology will be sketched. Afterwards, the planning and classroom practice of one statistics teacher will be outlined. Finally, the report stresses the structure of statistical knowledge and beliefs with regard to statistics of five of the teacher's students.

INTRODUCTION

In recent years, statistical reasoning (SR, Garfield, 2003), statistical literacy (SL, Gal, 2004) and statistical thinking (ST, Pfannkuch and Wild, 1999) have been declared as the three overarching goals of a modern statistics teaching. Since these three goals require a change of statistics teachers' instructional practice (e.g. Garfield, 2002), there is an increasing number of proposals from statistics educators as to how teachers can promote students' SR, SL and ST (e.g. Chance, 2002). In contrast, there are few research approaches that focus on SR, SL or ST concerning the classroom practice of "conventional" mathematics teachers (e.g. Watson, 2006).

The latter type of research approach is the basis for a qualitative research project that focuses on teachers' planning of statistics instruction (*teachers' individual curricula*), teachers' classroom practice (*teachers' factual curricula*), and the knowledge and beliefs students attained after statistics courses (*students' implemented curricula*). The motivation for the research project is the assumption that the nature of mathematics teachers' thinking is the key factor in any movement towards changing mathematics teaching (Chapman, 1999). Further, there is strong evidence that the knowledge and beliefs students attain are determined by their teachers' beliefs and their teachers' instructional practice (Calderhead, 1996). In this report a part of the larger research approach will be discussed, i.e. the individual curriculum and the factual curriculum of one statistics teacher, and the implemented curricula of five of the teacher's students. The report stresses the students' implemented curricula concerning the structure of the students' knowledge and beliefs regarding statistical concepts.

THEORETICAL FRAMEWORK

The distinction between three levels of a curriculum, i.e. the teachers' individual curricula, the teachers' factual curricula, and the students' implemented curricula is oriented to a model developed by Vollstädt, Tillmann, Rauin, Höhmann and Terbrügge (1999). Teachers' individual curricula, teachers' factual curricula, and students' implemented curricula are understood as part of an *action* in a psychological sense, which Erickson (1986, p. 126) defines as the "the physical behavior plus the meaning interpretations held by the actor".

In order to describe and structure the three levels of the curriculum, the psychological construct of *subjective theories* (Groben, Scheele, Schlee and Wahl, 1988) is used. Subjective theories are defined as a complex system of cognitions containing a rationale which is, at least, implicit. Hence, individual cognitions are connected in an argumentative mode. Subjective theories contain subjective concepts, subjective definitions of these concepts, and relations between these concepts that constitute the argumentative structure of the system of cognitions.

The teachers' individual curricula are understood as an argumentative system (subjective theory) comprising instructional contents, and instructional goals linked with these contents (Eichler, 2006). The teachers' instructional planning, i.e. the teachers' individual curricula, are non-observable intentions of action, which need to be reconstructed qualitatively by interpretation.

A teacher's factual curriculum is the observable part of his curriculum. It gives evidence for the appropriateness of the reconstruction of the teacher's individual curriculum. In other words, the teacher's factual curriculum provides evidence, if the teacher actually do what they say they intend to do.

The students' implemented curricula, i.e. the statistical knowledge and beliefs students attain due to the classroom practice, can be understood as a subjective peculiarity of SL, SR or ST. To structure the students' knowledge the construct of *statistical knowledge* will be used, which Broers (2006) describes as the core of SL, SR and ST. Furthermore Broers' distinction of declarative knowledge, procedural knowledge, and conceptual knowledge is used following the description of these three aspects of knowledge proposed by Hiebert and Carpenter (1992). In addition to the knowledge, the students' implemented curricula comprise beliefs concerning statistics or mathematics (Broers, 2006). As well as the teachers' individual curricula, the students' implemented curricula are non-observable and need to be reconstructed qualitatively by interpretation. The teachers' individual curricula comprise the teachers' conviction that their instructional practice yields their intended goals. Hence, the students' implemented curricula may yield evidence for the appropriateness of the teachers' convictions concerning their instructional goals.

METHODOLOGY

The methodology is based on case studies. One case is defined as the individual teacher plus five of his students. Cases are selected according to theoretical sampling. Data were collected with semi-structured interviews comprising several clusters of questions concerning the subjects shown in table 1.

An in-depth interview with one of the teachers took about two hours. The students were interviewed (about 30 minutes) one week after the teachers finished their statistics course. Interpreting transcribed interviews adheres to the principles of classical hermeneutics (Danner, 1998). The objective of this first phase of reconstruction is to identify subjective concepts and to see how they are defined. The second phase concerns the construction of argumentative systems of knowledge and beliefs, i.e. teachers' individual curricula or rather students' implemented curricula.

Interview with the teachers	**Interview with the students**
Contents of instruction	Stochastic concepts
Goals of mathematics instruction	Uses of mathematics instruction
The nature of (school) mathematics	The nature of mathematics
Teaching and learning mathematics	Teaching and learning mathematics
Institutional boundaries	Students' self-efficacy

Table 1: Clusters of the semi-structured interview

The classroom practice of the teachers was observed and recorded (in writing), for about four months during the time they taught statistics. Observing the teachers' factual curricula facilitates both evaluating the reconstruction of the teachers' individual curricula, and preparing prompts for the interviews with the students.

THE INDIVIDUAL AND FACTUAL CURRICULUM OF MR. D

Mr. D teaches stochastics in a special course of mathematics in grade 12 at a "Gymnasium" (a secondary high school). Many researchers (e.g. Helmke, 2007) provide criteria of a "good" classroom practice. Judging by these criteria, Mr. D can be described as a "good" teacher. Also, the students of Mr. D rate him as "good" teacher. His individual curriculum and factual curriculum will be outlined in brief.

Mr. D's curriculum concerning the instructional contents is traditional in Germany. The structure of the contents (reconstructed as individual curriculum and observed in the classroom practice of Mr. D) is shown in figure 1.

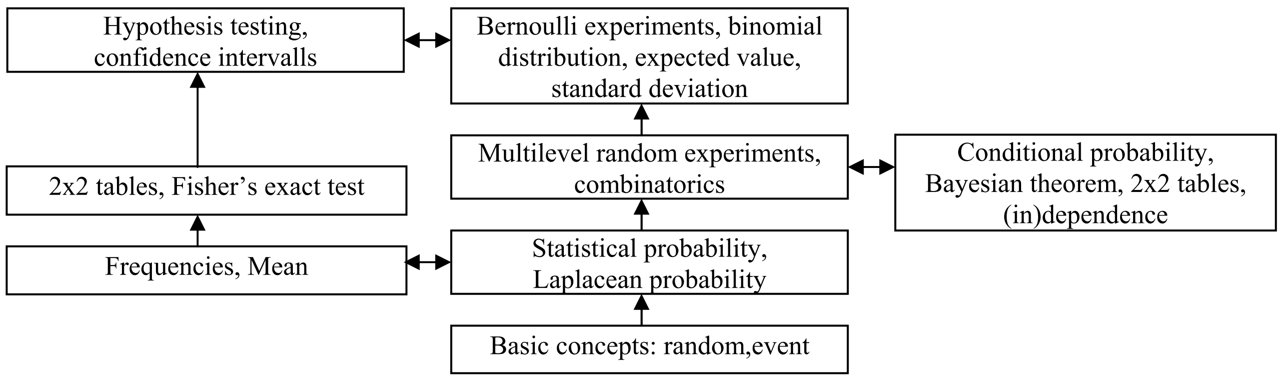


Figure 1: Instructional contents in the curriculum of Mr. D

Mr. D's curriculum concerning probability theory includes the concepts of chance, random experiments, probability, combinatorics, and binomial distribution. Mr. D primarily uses the statistical approach to probability, and teaches Laplacean probability as a second possibility. Besides this traditional curriculum, Mr. D teaches the concepts of conditional probabilitiy, of Bayesian theorem, and of (in)dependence as a digression from the main subject matter. This means that Mr. D spent a lot of time dealing with these concepts. However, these contents have no function concerning the other contents of Mr. D’s curriculum.

Mr. D wishes to develop statistical methods while teaching applications. The central goal of Mr. D is to develop these methods in a process, the result of which will be both the possibility to cope with real stochastical problems and the ability to criticise. Likewise, the central goal of the mathematics curriculum is to prepare students to cope with real-life mathematical problems. To attain these goals, Mr. D poses identical realistic problems at several times in his statistics course. Hence, Mr. D’s students examine one realistic problem with several statistical methods. For example, one of Mr. D’s central and recurrent problems is concerned with elections.

Another goal of Mr. D is to establish a network of mathematical or statistical concepts. Regarding this goal, there is a break between Mr. D’s individual curriculum and Mr. D’s factual curriculum. So, Mr. D seldom shows relationships between statistical concepts explicitly, and he never uses concept-maps or similar strategies to emphasise the idea of a network of statistical concepts.

THE IMPLEMENTED CURRICULA OF FIVE OF MR D'S STUDENTS

The discussion of the implemented curricula of five of Mr. D’s students comprises aspects like statistical knowledge and beliefs about the relevance of statistics. The discussion starts with Friederike. Additionally, the knowledge and beliefs of the other four students will be mentioned.

Statistical knowledge

According to Mr. D, Friederike is a (female) student with high aptitude. Friederike’s statistical knowledge is shown in figure 2. The single statistical concepts (bold font) are understood as the students *declarative knowledge*. Friederike remembers the quoted concepts and is able to explain these concepts (except those in double brackets, e.g. conditional probability). The four concepts on the right side (italics), e.g. probability trees, are understood as the *procedural knowledge* of Friederike. The *conceptual knowledge* of Friederike is represented in three ways:

1. The clusters of concepts represent the students knowledge concerning relationships between statistical concepts within a representation form (Hiebert and Carpenter 1992), e.g. the relationship between the concepts Bernoulli experiment, binomial distribution and random variable:

   Friederike: As far as the Bernoulli experiments are concerned, they can be calculated by using binomial distribution. And we can do this for certain values of X, which is the random variable.

2. Another aspect of the conceptual knowledge comprises the connections of different representation forms. Such relationships exist between statistical concepts and applications of the statistical concepts (in angled brackets), e.g. the relationship between the concept of hypothesis testing and the application of hypothesis testing, for example the extrapolation concerning elections. Most of the applications of the statistical concepts emanate from the classroom practice of Mr. D (if not, the examples are underlined).
3. The third aspect of the conceptual knowledge is the relationship between several clusters of statistical concepts, which is represented by arrows. The dashed arrows are used if the relationship is vague.

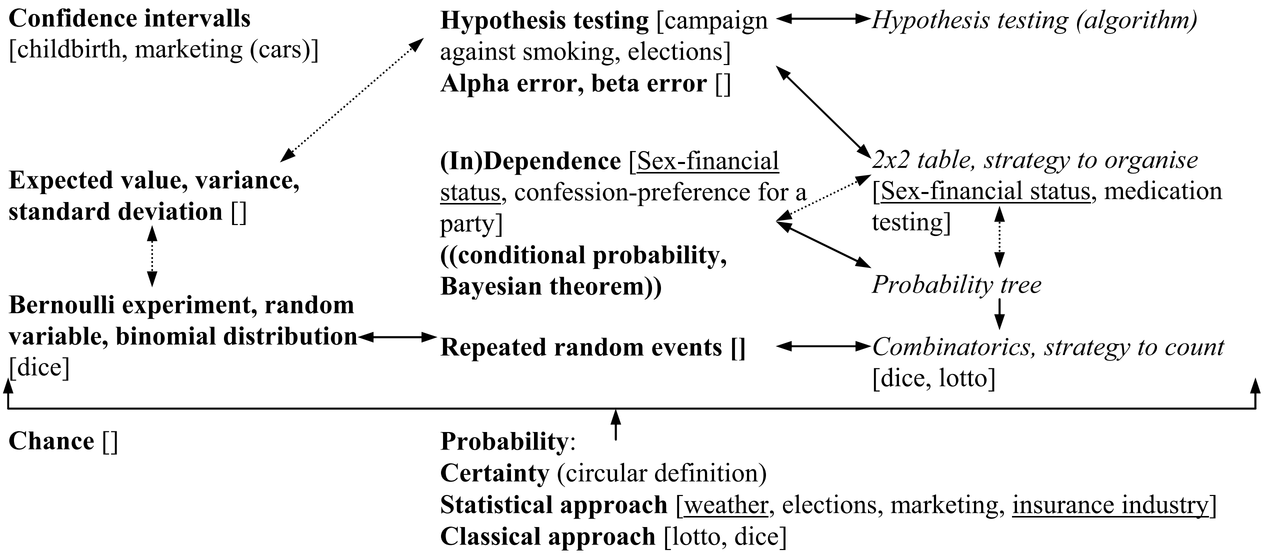


Figure 2: Statistical knowledge of Friederike

One example concerning the third aspect of the conceptual knowledge is illustrated by the relationships between the statistical concepts of (in)dependence and the statistical procedures of the probability tree and the 2x2 table. Friederike remembers the 2x2 table as a strategy to handle the Fisher exact test:

> Friederike: Yes, a 2x2 table, it’s much easier to grasp. At first we had that with much smaller numbers, for example medical tests, and then we simply transposed it into the 2x2 table.

Although Friederike remembers the relation between the 2x2 table and the Fisher exact test, she is not able to explain the algorithm of the Fisher exact test. Instead, she uses the 2x2 table in a descriptive way, mentioning an example that represents the subject matter of (in)dependence, i.e. the relationship between sex and financial status:

> Friederike: So, for example, you could calculate the relative frequency of something, for example the relative frequency of women or men, and the relative frequency with which a certain criterion [rich or poor] applies to them.

The relationship between the concept of (in)dependence and the procedure of the 2x2 table is vague because to explain the concept of (in)dependence, Friederike uses only the procedure of the probability tree:

Friederike: So, if something is independent, then the branches on the tree, their probabilities are the same [...] and if it's dependent, then there are different probabilities, as with male and female, and then the probabilities that follow on are not the same.

Regarding the declarative knowledge, Friederike is able to explain most of the statistical concepts that are contained in the factual curriculum of Mr. D, except for the statistical concepts which Mr. D teaches as a relatively long digression, i.e. conditional probability and the theorem of Bayes. The lack of knowledge regarding these two statistical concepts is symptomatic for the students of Mr. D. Regarding all the statistical concepts, Friederike's declarative knowledge is wider than the declarative knowledge of the other four students of Mr. D (see table 2):

Student	**Concepts that the students could *not* remember or could *not* explain**
Gina	(in)dependence, conditional probability, Bayesian theorem, binomial distribution
Hans	(in)dependence, conditional probability, Bayesian theorem, random variable, combinatorics
Ingo	Conditional probability, Bayesian theorem
Janine	conditional probability, Bayesian theorem, random variable, Bernoulli experiment, binomial distribution, combinatorics, alpha error, beta error

Table 2: Declarative knowledge of Mr. D's students

Regarding the procedural knowledge, Friederike is able to handle graphical procedures like the probability tree or the 2x2 table. However, she is only able to use the tree in an appropriate way and not the other. Furthermore, Friederike explains the algorithm of hypothesis testing in a sophisticated way. Although Friederike is able to explain most of the statistical concepts, she is not able to mention formulas, expect the formula of the expected value concerning the binomial distribution (n times p). Regarding all five students of Mr. D, it is symptomatic that the students are able (1) to use only the probability tree in an appropriate way, (2) to use the 2x2 table to tabulate frequencies (probabilities), (3) to explain the algorithm of hypothesis testing. Finally, the students are not able to remember formulas.

Regarding the conceptual knowledge, Friederike has only a vague idea of how to connect statistical concepts. Although she is able to perceive both the relationships concerning several clusters of statistical concepts, and the relationships between statistical concepts and their applications, the clusters of statistical concepts are often unrelated. Friederike remembers the clusters of statistical concepts as isolated book chapters. For example, she explains the cluster Bernoulli experiment – binomial distribution – random variable. However, she does not perceive the relation between this cluster and the concept of hypothesis testing, which is based on the binomial distribution in the factual curriculum of Mr. D. Nevertheless, Friederike's network of knowledge has more interconnections then the conceptual knowledge of the other students. The following table shows the relationship between concept clusters of the students of Mr. D, except for the concept of probability, which is interconnected to all the other statistical concepts in the network of knowledge that all students have.

Student	Relationship between concept clusters
Gina	(1) Expected value, standard deviation → hypothesis testing ← algorithm of hypothesis testing (2) Bernoulli-Experiment → Expected value, standard deviation (vague) (3) Probability tree → Repeated random events → Bernoulli-Experiment (4) 2x2 table → hypothesis testing
Hans	(1) algorithm of hypothesis testing → hypothesis testing (2) Repeated random events → Bernoulli-Experiment (3) 2x2 table → hypothesis testing
Ingo	(1) algorithm of hypothesis testing → hypothesis testing (2) combinatorics → Repeated random events → Bernoulli-Experiment (3) Probability tree → (in)dependence (4) 2x2 table → hypothesis testing
Janine	(1) Expected value, standard deviation → hypothesis testing ← algorithm of hypothesis testing (2) Probability tree → combinatorics (vague)

Table 3: Conceptual knowledge of Mr. D's students

Beliefs about the relevance of statistics

The structure of Friederike's beliefs about the relevance of statistics (and mathematics) is shown in figure 3.

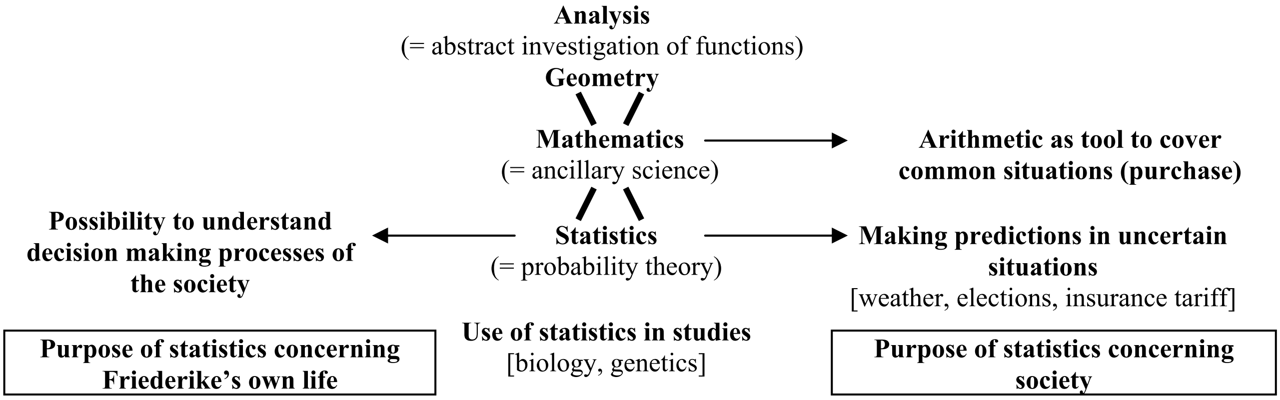


Figure 3: Friederike's beliefs about the relevance of statistics (and mathematics)

The relations used to describe the students' beliefs have the following meaning: The "/ \" describes the hierarchy of concepts (which are subjectively defined in brackets). For example, for Friederike, mathematics is the superordinate concept for statistics or analysis. The second relation, i.e. the arrow, describes the purpose of the mathematical disciplines. Finally, there is the distinction between the purpose of mathematics concerning the students' own life and concerning the society.

Friederike believes that statistics are a tool to solve real problems emerging in the society. Real problems, for example, exist in predicting events in uncertain situations. Friederike mentions, as one example (in angled brackets) for this purpose, the construction of an insurance tariff:

Friederike: So, I think, for example, that you can use this in the insurance business. That is, in all the areas where you need a certain amount of accuracy but you can never be 100 percent certain. Yes, in insurance, where some things that are very probable require a premium that is very high. And other things, which are not so probable, are not such a great risk. They would rarely have to pay a compensation.

For Friederike, a further purpose of statistics is their use in a course of university studies. Friederike mentions biology and in particular genetics as an example. She is, however, not able to explain how statistics may be a tool to solve problems concerning biology. In the same way, Friederike's beliefs about the relevance of statistics for her own life are vague. She mentions the role of statistics in understanding specific decision-making processes. She is, however, not able to explain this role with regard to a concrete example. Finally, Friederike says that mathematics offers little more than arithmetic:

Friederike: Outside school, maths has no real significance. Ok, you can deal with numbers, but we learned how to do that a long time ago..

Interviewer: Ok, and what you do later on in school, isn't that important?

Friederike: Perhaps with stochastics. Perhaps you can better understand how probable something might be.

All the students of Mr. D define statistics as the theory of probabilities. Most of the students explain the purpose of statistics for the society by using examples from the classroom instruction (see table 4). With the exception of Hans, the students are not able to explain the benefit of statistics for their own lives. Finally, most of the students argue that the benefit of mathematics is to learn arithmetic, which enables someone to deal successfully with money (shopping etc.).

Student	Purpose of statistics (societey/own life)	Purpose of mathematics (society/own life)
Gina	prediction (e.g.medical science) / games (unimportant)	shopping / career
Hans	none / estimation regarding games (casino)	none / economics, career
Ingo	prediction (e.g.elections) / games (unimportant)	shopping / programming, study, career
Janine	economics / critical faculty (press coverage)	shopping / none

Table 4: Friederike's beliefs about the relevance of statistics (and mathematics)

DISCUSSION

Regarding the declarative knowledge of the five students it is a crucial result that all students have no or little knowledge concerning the concepts of conditional probabilities, of (in)dependence, or of the Bayesian theorem. To facilitate the students' comprehension of independence and the Bayesian theorem, Mr. D uses the probability tree in a traditional manner. It is possible that the students would have more knowledge concerning this concept that have an intrinsic difficulty, if Mr. D uses the tree with natural frequencies in a way many authors suggested (e.g.

Martignon and Wassner 2002). It seems, however, that students predominantly are not able to remember or to explain the concepts of an instructional digression. Regarding the procedural knowledge it is remarkable that the five students are able to set up probability trees or 2x2 tables. However, none of the students is able to explain how to use the 2x2 table to solve specific problems except for arranging frequencies. It seems that Mr. D's decision to use the 2x2 table concerning two very different statistical concepts (Fisher exact test, dependence) tends to impede the students' comprehension.

Although it is one primary goal of Mr. D to establish a network of statistical concepts, the conceptual knowledge of his students is limited. The students remember the clusters of statistical concepts as isolated book chapters. On this note, most of the students have knowledge about the relationship of the concepts of the Bernoulli experiment, of binomial distribution and of a random variable. The students, however, do not mention the relationship between the concepts of binomial distribution and independence or binomial distribution and hypothesis testing.

The students mention several examples of applying statistics in real world problems linked to the statistical concepts. Most of these examples emanate from Mr. D's instruction. However, although Mr. D emphasises both perceiving, and coping with real statistical problems, his students do not understand the purpose of statistics in dealing with problems of the real world (with no regard to instructional examples).

Above, Mr. D was rated above as "good" teacher in respect to his classroom practice. However, the realised curricula of five of his students show a phenomenon that Helmke (2007) describes as follows: a teacher's well prepared classroom practice must not necessarily yield the impact the teacher intends.

OUTLOOK

The purpose of this report was to discuss the impact of one teacher's everyday classroom practice on students' statistical knowledge in a descriptive way. The case of Mr. D and his five students is one of four cases of the larger research project. The aim of this ongoing project will be to understand the relationship between the teachers individual curricula, the teachers factual curricula and the students implemented curricula. The aim of this holistic understanding will be to find starting points to change the teachers' everyday classroom practice regarding both the teacher's individual curricula and the demands concerning the promotion of SR, SL and ST.

REFERENCES

Broers, N. J.: 2006, Learning Goals: The Primacy of Statistical Knowledge. In Rossman, A. and B. Chance, editors, *ICOTS-7 Conference Proceedings*. IASE, Salvador (CD-Rom).

Calderhead, J.: 1996, Beliefs and Knowledge. In Berliner, D. C., editor, *Handbook of Educational Psychology*. Macmillan, New York.

Chance, B. L.: 2002, Components of Statistical Thinking and Implications for Instruction and Assessment. *Journal of Statistics Education* 10(3).

Chapman, O.: 1999, Researching Mathematics Teacher Thinking. In Zaslavsky, O., editor, *Proceedings PME 23*, Volume 2, pp. 385-392. PME, Haifa.

Danner, H.: 1998, *Methoden geisteswissenschaftlicher Pädagogik*. Tübingen, Mohr.

Eichler, A. 2006, Individual Curricula – Beliefs behind Beliefs. In Rossman, A. and B. Chance, editors, *ICOTS-7 Conference Proceedings*. IASE, Salvador (CD-Rom).

Erickson, F.: 1986, Qualitative Methods in Research on Teaching. In Wittrock, M. C., editor, *Handbook of Research on Teaching*, pp. 119-161. Macmillan, New York.

Gal, I.: 2004, Statistical Literacy: Meanings, Components, Responsibilities. In Ben-Zvi, D. and J. Garfield, editors, *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*. Kluwer, Dordrecht.

Garfield, J.: 2002, The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education* 10(3).

Garfield, J.: 2003, Assessing statistical reasoning. *Statistics Education Research Journal* 2(1).

Groeben, N., Scheele, B., Schlee, J. and D. Wahl: 1988, *Forschungsprogramm Subjektive Theorien*. Franke, Tübingen.

Helmke, A.: 2007, Unterrrichtsqualität. Kallmeyer, Seelze.

Hiebert, J. and T. P. Carpenter: 1992, Learning and Teaching with Understanding. In Grouws, D. A., editor, *Handbook of Research on Mathematics Teaching and Learning*, pp. 65-97. Macmillan, New York.

Martignon, L. and C. Wassner: Teaching Decision Making and Statistical Thinking with Natural Frequencies. In Phillips, B., editor, *ICOTS-6 Conference Proceedings*. IASE, Cape Town (CD-ROM).

Pfannkuch, M. and C. Wild: 1999, Statistical Thinking in Empirical Enquiry. *International Statistical Review* 67(3), 223-248.

Vollstädt, W., Tillmann, K.-J., Rauin, U., Höhmann, K. and A. Tebrügge: 1999, *Lehrpläne im Schulalltag*. Leske and Budrich, Opladen.

Watson, J. M.: 2006, Issues for Statistical Literacy in the Middle School. In Rossman, A. and B. Chance, editors, *ICOTS-7 Conference Proceedings*. IASE, Salvador (CD-Rom).

THE 'SAME' PROBLEM IN THREE PRESENTATION FORMATS: DIFFERENT PERCENTAGES OF SUCCESS AND THINKING PROCESSES

M. Pedro Huerta (1) Mª Ángeles Lonjedo (2)

(1) Universidad de València, Spain, (2) IES Montserrat, Spain

ABSTRACT

In this paper we investigate the influence that presentation format of a conditional probability problem has on students' problem solving behavior. We not only focus on the way numerical data is presented but also on information in text form that refers to a conditional probability. We report that students' behavior changes depending on data presentation, and the percentage of students that succeed in solving a problem increases if we change the presentation format of the problem in a suitable way. We conclude by making some proposals for teaching conditional probability problem solving.

INTRODUCTION

From a psychological point of view, some authors (eg, Gigerenzer & Hoffrage, 1995; Cosmides & Tooby, 1996; as opposed to Evans, Handley; Perham, Over & Thompson, 2000; Girotto & Gonzalez, 2001), suggest that people calculate Bayesian inferences better if the information is expressed in terms of frequencies rather than probabilities. They report that participants perform better because there is a strong relationship between data format and the rules required to answer the problem. In (natural) frequency formats, these rules are less complex than in probability formats and can facilitate reasoning in complex Bayesian situations. Other authors argue that the reason doesn't lie in data format but in information structure and the form of the question, making the problem much easier to understand.

Almost all the conditional probability problems used in the research mentioned above were structurally isomorphic. Usually these problems were considered in pairs. Apart from data format, all of them can be mathematically (or symbolically) read as follows: Known $p(E)$, $p(+|E)$ and $p(+|\sim E)$ calculates $p(E|+)$, otherwise known as the Disease Problem. All of these problems we label by means of a vector (1,0,2) (see Carles & Huerta, 2007) that indicates the number and type of known numerical data we have in text of problem: in this case, we use one absolute probability and two conditional probabilities to calculate the unknown numerical data. Number 0 means that we do not know the intersection probabilities. In order to calculate $p(E|+)$ we need Bayes' rule and the Theorem of Total Probability. But these problems are different (ie, not isomorphic) if we consider that problem information is presented using different data formats: frequencies in one case and percentages in the other. None of them used probabilities (numbers in [0,1]) as a data format.

On the other hand, the subjects performing problems in these studies weren't considered to be math students, with some knowledge of making Bayesian inferences, but as people naïve in this topic and in probabilistic reasoning. For this reason, the main objective of these investigations was to explore reasons for successful problem solving according to data format and discovering which data format facilitates the most success.

From the point of view of the didactic of mathematics, in a previous paper (Huerta & Lonjedo, 2006) we showed that different presentation formats of a conditional probability problem resulted in different student problem solving behavior. This paper investigated the processes involved in solving 16 conditional probability problems. These paired problems were structurally isomorphic but used different data formats. One of the conclusions relates to differences in student behavior in solving the 'same' problem when data is expressed in terms of percentages as opposed to probabilities; ie, if data is expressed in terms of percentages, then students usually solve these problems using mainly arithmetical thinking strategies, whereas if data is expressed in terms of probabilities, one can recognize probabilistic thinking strategies in solving these problems.

In CERME4 (Huerta & Lonjedo, 2006) we presented a report highlighting the problem solving processes when data is expressed in terms of percentages and probabilities. In CERME5, however, we present a piece of work that focuses on the process of solving three problems that are isomorphic in structure but where the information is expressed in three different formats: in terms of percentages, probabilities and absolute frequencies. Thus, the aims of this paper are: (1) to study the influence of data format in conditional probability problems on students' behaviors and success; (2) to study the influence of semantic and syntactic aspects on the students' success in solving these problems.

THE RESEARCH PROBLEM

Let us consider problems P7 and P15 (see Table 1) used in Huerta & Lonjedo (2006). These problems can be mathematically read as follows: Known $p(A)$, $p(B)$ and $p(A|B)$, calculate $p(B|A)$. It is a (2,0,1) problem asking for an inverse probability from a known conditional probability. The Bayes' rule solves this problem.

We know that in general, the percentage of students that were successful in solving both problems was very low (Huerta & Lonjedo, 2006). Only math college's students were successful whereas secondary school students were not. Therefore, we considered a new problem, P1, structurally isomorphic with both P7 and P15, but with the data expressed in a different format. From the experience we had with problems P7 and P15, we also reconsidered the information in text form that was used to describe the data as conditional probabilities and expressed them with the same grammatical structure, in order to avoid as many misunderstandings as possible. One of these misunderstandings, for example, relates to students' confusion between the conditional probability and the intersection probability. Quantities in P1 are always absolute frequencies, except when referring to a conditional probability, in which case percentages must be used. Moreover, in P1 we also tried to avoid both semantic difficulties, using for example a more understandable sentence for the conditional probability, and semiotic difficulties, writing data for the absolute probabilities in terms of absolute (natural) frequencies. In Table 1 we can see the three problems we are referring to which are, essentially, three versions of the 'same' problem.

P7 PROBLEM Data in percentages	*60% of students in a school succeeded in Philosophy and 70% in Mathematics. Moreover, 80% of the students that succeeded in Mathematics also succeeded in Philosophy. If Juan succeeded in Philosophy, what is the probability that he also succeeded in Mathematics?*
P15 PROBLEM Data in probabilities	*In a school, the probability of success in Philosophy is 0.6 and in Mathematics, 0.7. Choosing a student at random among those that succeeded in Mathematics, the probability that he/she also succeeded in Philosophy is 0.8. If Juan succeeded in Philosophy, what is the probability that he also succeeded in Mathematics?*
P1 PROBLEM Data in frequencies	*In a class of 100 students, 60 succeeded in Philosophy and 70 succeeded in Mathematics. Among those who succeeded in Mathematics, 80% also succeeded in philosophy. Of those who succeeded in Philosophy, what percentage of students also succeeded in Mathematics?*

Table 1. Three versions of the 'same' problem

These problems form part of a broader research that tries to investigate the processes involved in solving conditional probability problems. One of the questions we try to answer has to do with the relationship between data presentation format and students' problem solving process.

METHOD

All three problems were items in a test administered to students of different ages and mathematical ability: Lower secondary school (13-14 year olds), upper secondary school (15-18 year olds) and 2nd year math students at university. In table 2 we can see the distribution of the student sample involved in this research. Only students from upper secondary school and university were taught about conditional probability, whereas students from lower secondary school were not.

School Level	P7	P15	P1
Lower Secondary School	11	5	31
Upper Secondary School	52	26	39
University (Math College)	4	2	10
Total	67	33	80

Table 2: Number of students that tried to solve each problem

Because we were not only interested in students' success in solving problems but also in resolution processes, successful or not, we designed a set of descriptors to analyze problem solving behaviour, as follows:

1. - Problem solving process with success. This descriptor reports students' successful behaviors in finding the correct result to the problem. Depending on the different reasoning shown by students during the problem solving process, we distinguish:

1.1. Problem solving processes that include a type of thinking that is exclusively arithmetical: Students think in quantities and not in events and their probabilities, at least in a conscious way.

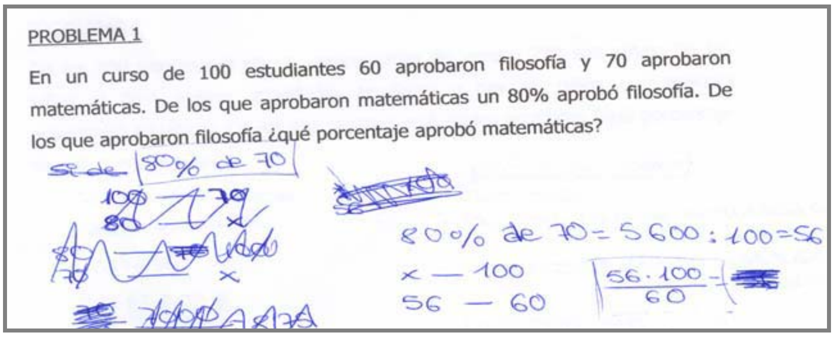
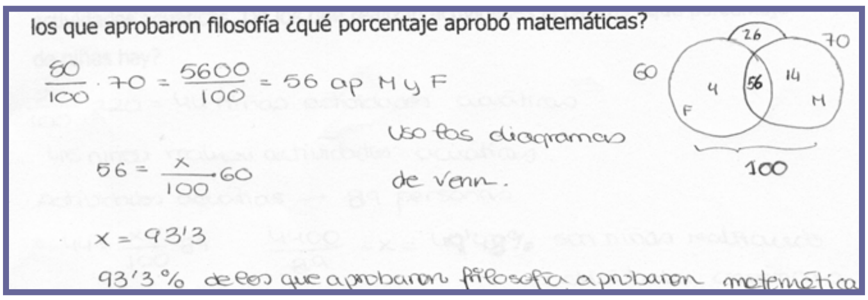


Figure 1: Example of thinking process (in P1) classified in 1.1, exclusively arithmetical thinking.

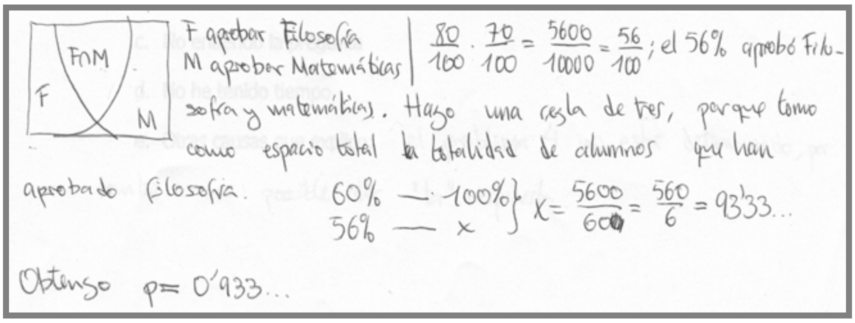
1.2. Problem solving processes that include a type of thinking that is mostly arithmetical: Students think in quantities but they recognize events and their associated frequencies or percentages.



F to succeed in Philosophy, M to succeed in Mathematics. 56 succeeded in M and F, I use the Venn diagrams, 93.3% among those successful in Philosophy succeeded in Mathematics

Figure 2: Example of thinking process (in P1) classified in 1.2, mostly arithmetical thinking.

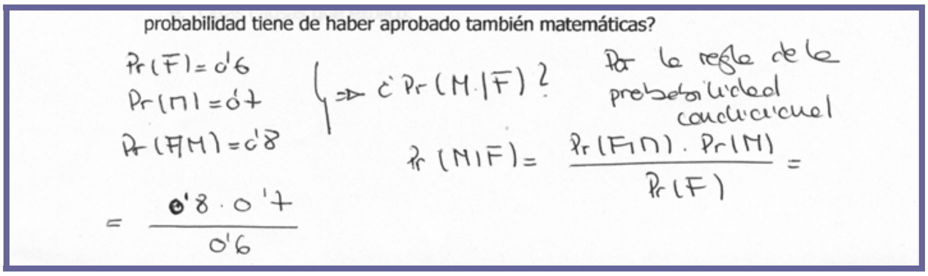
1.3. Problem solving processes that include a type of thinking that is basically probabilistic. In solving problems, students think arithmetically in quantities. These quantities are not used explicitly as probabilities. However, students recognize events and assign probabilities to them without using probability rules in a conscious way.


F∩M
F
M
F aprobar Filosofía
M aprobar Matemáticas
$\frac{80}{100} \cdot \frac{70}{100} = \frac{5600}{10000} = \frac{56}{100}$; el 56% aprobó Filosofía y matemáticas. Hago una regla de tres, porque tomo como espacio total la totalidad de alumnos que han aprobado filosofía.
$\left.\begin{array}{l} 60\% \text{ — } 100\% \\ 56\% \text{ — } x \end{array}\right\} x = \frac{5600}{60} = \frac{560}{6} = 93'33\ldots$
Obtengo p= 0'933...

F to succeed in Philosophy, M to succeed in Mathematics … 56% succeeded in Philosophy and Mathematics. I do a rule of three because I get as total space the totality of students that succeeded in Philosophy ….I get p= 0.933..

Figure 3: Example of thinking process (in P7) classified in 1.3, basically probabilistic thinking.

1.4 Problem solving processes that include a type of thinking that is exclusively probabilistic. Students recognize events, assign probabilities to the events and explicitly use probability rules in order to find the result of the problem.


probabilidad tiene de haber aprobado también matemáticas?
$\left.\begin{array}{l} Pr(F)=0'6 \\ Pr(M)=0'7 \\ Pr(F\cap M)=0'8 \end{array}\right\} \Rightarrow ¿Pr(M|F)?$
Por la regla de la probabilidad condicional
$Pr(M|F) = \frac{Pr(F\cap M) \cdot Pr(M)}{Pr(F)} = \frac{0'8 \cdot 0'7}{0'6}$

Because the rule of conditional probability

Figure 4: Example of thinking process (in P15) classified in 1.4, exclusively probabilistic thinking.

2. - Problem solving processes without success. This descriptor reports students' behaviors that were unsuccessful. Within this general descriptor, we consider a more specific descriptor that describes students' semantic and syntactic difficulties, misunderstandings and mistakes, as follows:

2.1 Difficulties. We analyze the solvers' difficulties related to semantic and semiotic variables.

2.1.1 Semantic Difficulties. We analyze grammatical structures in descriptions used to express conditionality both as data (known and unknown) and as text.

2.1.2 Syntactic Difficulties. We analyze formats of data and question presentation in problems.

2.2 Mistakes. We analyze students' mistakes related to difficulties.

2.2.1. Mistakes as a result of semantic difficulties. These mistakes are undesirable interpretations of data in problems when students are translating them from usual language into symbolic language. Sometimes problem solving processes are coherent with students' interpretations. These mistakes could appear early in the process, both in recognizing events and in assigning probabilities to events.

2.2.1.1. Students' interpretations of conditional probability when it is data. We can distinguish the following interpretations:

2.2.1.1.1. Interpretation of the conditionality as an intersection event (Ojeda, 1995).

2.2.1.1.2. Interpretation of the conditionality as an absolute probability. Student answers the question $p(A|B)$ by means of $p(A)$.

2.2.1.2. Students' interpretations of the conditional probability when it is a question.

2.2.1.2.1. Interpretation of the conditionality as an intersection event. Students answer questions about a conditional probability by means of an intersection probability.

2.2.1.2.2. Interpretation of the conditionality in question as conditionality in data. Student interprets $p(A|B)$ as equal to $p(B|A)$, the first being probability in data and the second, the question.

3. - Others. We place here all students' resolutions that are impossible to be qualified or explained by the other descriptors. These include blank answers, answers without workings, unrecognizable signs etc.

There is another source of mistakes in solving these problems that we did not consider in this work. This concerns the misuse of decimal numbers, percentages, formulas and mathematical calculations. These mistakes, of course, would hinder successful problem resolution.

SOME RESULTS

The percentage of students that succeeded with problems P7 and P15 was 6%. All of them were students from University. However, the percentage of students that succeeded with P1 was 36.25%. This figure includes students from all age levels.

From the data in Table 3, we can see that the percentage of students that did not try to solve the problems decreases in order, from the highest (66.67%) when the data is in the form of probabilities, to the lowest (21.25%) when it is presented as frequencies. Consequently, there is an appreciable increase in the percentage of students that succeeded in solving the problems, from the lowest when the data is in the form of probabilities and percentages (5.97% and 6.06% respectively), to the highest (36.25%) when presented as frequencies. In other words, the success rate when data was presented as percentages or probabilties was extremely low compared with the number of students that succeeded with the data in the form of frequencies. We can explain these differences using the descriptors that classify students' mistakes in the problem solving process. There is a very high percentage (89.74%) of students that did not succeed in P7 because they incorrectly interpreted conditionality data either as a question or as known data. Similar mistakes occurred in P15. However, this misunderstanding occured much less (20.6%) with data in the form of frequencies. A very high percentage (89.65%) of students that were successful when the problem was described using frequencies used arithmetical reasoning. No one used this type of reasoning when problem solving with percentages or probabilities. Only 3 students out of the 29 that succeeded with the problem in frequencies used probabilistic reasoning.

DISCUSSION

The presentation of data in a conditional probability problem has some influence on the students' success and problem solving behavior (see Table 3). We think that the increase in students' success (Table 3) is due to two factors: avoiding words that provoke ambiguity; and presenting data as absolute frequencies. For example, avoiding words like *y* (and) or *también* (also) in problem descriptions prevented students from confusing conditional probability with intersection probability. The expression *De los que* (Among those who), that refers to conditionality both as data and as question, improved students' interpretation.

When data is expressed in terms of absolute frequencies and conditional probability as a percentage (Lonjedo Huerta (2006), p. 531), we believe that the chances of successful problem resolution are enhanced with a consequent increase in the percentage of successful students. Gigerenzer (1994) reports that in order to solve probability problems, our minds are better equipped if all data is expressed in terms of frequencies. We agree, although when referring to a conditional probability problem, we would like to add that if one of the data is a conditional probability, then it must be expressed in terms of a percentage in order to differentiate it from other data expressed in terms of absolute frequencies. In this manner, we can help students to interpret conditionality correctly.

Descriptors	P7 Problem Data in percentages	P15 Problem Data in probabilities	P1 Problem Data in frequencies
1. Problem solving process with success	(4 out of 67) 5.97%	(2 out of 33) 6.06%	(29 out of 80) 36.25%
1.1 Thinking is exclusively arithmetical	0	0	(7/29) 24.13%
1.2 Thinking is mostly arithmetical	0	0	(19/29) 65.52%
1.3 Thinking is basically probabilistic	(1/4) 25%	0	0
1.4 Thinking is exclusively probabilistic	(3/4) 75%	(2/2) 100%	(3/29) 10.34%
2. Problem solving processes without success	(39 out of 67) 58.21%	(9 out of 33) 27.27%	(34 out of 80) 42.50%
2.2.1.1.1. Interpretation of the conditionality as an intersection event	(13/39) 33.33%	(4/9) 44.44%	(1/34) 2.94%,
2.2.1.1.2. Interpretation of the conditionality as an absolute probability	0	(1/9) 11.11%,	0
2.2.1.2.1. Interpretation of the conditionality in question as an intersection event	(22/39) 56.41%	(4/9) 44.44%	(6/34) 17.65%
2.2.1.2.2. Interpretation of the conditionality in question as conditionality in data	(3/39) 7.69%	0	(2/34) 5.8%,
3. Others	(24 out of 67) 35.82%	(22 out of 33) 66.67%	(17 out of 80) 21.25%

Table 3. Percentages of students in relation to the descriptors.

Of those students who succeeded, arithmetical thinking strategies were typically used in problem solving, although they also demonstrated recognition of events and their frequencies and percentages. However, the thinking process used in solving the problems seems to be strongly related to data format. Students that succeeded with the problem described in frequencies used mainly arithmetical reasoning, whereas those who were successful with the problem expressed as percentages or probabilities used probabilistic reasoning. In general, when students were solving these problems they use the data explicitly mentioned without translation from one format to another. Only in a few cases (3 out of 29) did students translate frequencies into probabilities in order to solve the problem (P1) using probabilistic reasoning. These were all University students.

CONCLUSION

The concept of natural frequencies introduced by Gigerenzer and his colleagues (a good discussion about this concept can be read in Hoffrage, Gigerenzer, Krauss & Martignon (2002)), has produced some proposals about natural frequencies-based teaching in the last two ICOTS. Martignon & Wassner (2002) proposed using (natural) frequency trees in order to read quantities in a typical Bayesian problem and facilitate secondary school students problem solving. They concluded that students trained with frequency trees performed significantly better than students formula-trained. In a similar way, but in a computer environment, Sedlmeir (2002) proposed using frequency trees and frequency grids to help students read (1,0,2) problems with data in percentages using an isomorphic problem with data in (natural) frequencies. Moreover, Martignon & Kurz-Milcke (2006) proposed a method of training younger students by means of arithmetic urns and tinker-cubes in stochastic reasoning.

This work also has implications for teaching problem solving in conditional probability. In agreement with the authors mentioned above, we propose organizing the process of teaching this topic by starting with solving problems like P1, prior to attempting problems like P7, and then finally moving on to problems like P15. Our reasoning consists of introducing students to the subject by means of rates and proportions, making Bayesian inferences with data in (natural) frequencies and using arithmetic reasoning, followed by percentages employing basically probabilistic reasoning, and finally by means of probabilities with an exclusively probabilistic reasoning approach. (0,0,3) problems, like the P4 problem that was presented by Carles & Huerta (2007) in CERME5, are not able to be solved using only arithmetic reasoning and, consequently, by means of (natural) frequencies. They require probabilities and probabilistic reasoning strategies.

Finally, we also agree with other authors (eg, Ojeda, 1995; Girotto & Gonzalez, 2001) that one of the main sources of student error involves misinterpretation of the conditionality and the probability of the intersection event. The incidence of this misinterpretation could be reduced if we teach conditional probability problem solving in the way we propose in this work.

REFERENCES

Carles, M.; Huerta, M. P. (2007). *Conditional probability problems and contexts: The Test of Diagnostic Context*. Communication presented to the

Stochastic Working Group at CERME V. Larnaca, Cyprus. February 2007.

Cosmides, L.; Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition* 58, pp. 1-73.

Evans J.St.B.T.; Simon, J.H.; Perham, D.E.O; Thompson, V.A. (2000). Frequency versus probability formats in statistical word problem. *Cognition* 77, pp. 197-213.

Gigerenzer, G. (1994): Why the distinction between single-event probabilities and frequencies is important for psychology (and vice-versa), In G. Wright y P. Ayton (Eds.) *Subjective probability*, pp. 129-161, Wiley

Gigerenzer, G.; Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency format. *Psychological Review*, 102, pp. 684-704.

Girotto, V.; Gonzalez, M. (2001). Solving probabilistic and statistical problems: a matter of information structure and question form. *Cognition* 78, pp. 247-276.

Hoffrage, U.; Gigerenzer, G.; Krauss, S.; Martignon, L. (2002). Representation facilitates reasoning: what natural frequencies are and what they are not. *Cognition*, 84, pp. 343-352.

Huerta, M.P; Lonjedo, M. A. (2006) The Nature of the quantities in a conditional probability problem. Its influence in the problem solving behavior. In M. Bosch (ed) (2006), *European Research in Mathematics Education IV*. *Proceedings of the Four Congress of the European Society for Research in Mathematics Education*, pp. 528-538. Sant Feliu de Guíxols, Spain.

Martignon, L., Wassner, Ch. (2002). *Teaching Decision Making and Statistical Thinking with Natural Frequencies*, in http://www.stat.auckland.ac.nz/publications/1/10_52_ma.pdf

Martignon, L.; Kurz-Milcke, E. (2006). *Educating Children in Stochastic Modeling: Games with Stochastic Urns and Colored Tinker-Cubes*, in http://www.stat.auckland.ac.nz/publications/17/C443.pdf

Ojeda Salazar, A.M. (1995). Dificultades del alumnado respecto a la probabilidad condicional, *UNO Revista de Didáctica de las Matemáticas*, nº5, julio 1995, pp. 37-44

Sedlmeier, P. (2002). *Improving Statistical Reasoning by Using the Right Representational Format*, in http://www.stat.auckland.ac.nz/publications/1/3f3_sedl.pdf

THE RELATIONSHIP BETWEEN LOCAL AND GLOBAL PERSPECTIVES ON RANDOMNESS [1]

Peter Johnston-Wilder and Dave Pratt

Institute of Education, University of Warwick, UK, and Institute of Education, University of London, UK.

In clinical interviews, learners were invited to talk about their experiences of making sense of the emerging sequence of outcomes from repeated trials using different generators, some of which were biased. Analysis of the interviews revealed distinct ways of viewing the phenomena represented by the interview tasks. Drawing upon the local and global meanings of randomness identified by Pratt (1998), learners were found to shift their attention rapidly between local and global perspectives. Data presented in this paper illustrates the shifting perspectives. It is suggested that not only did these students, three or more years older than those in Pratt's study, shift attention from global to local as well as from local to global perspectives, but such shifts may have been stimulated at times by consideration of causal factors.

STIMULUS TASKS

When studying perceptions of randomness, it is important to distinguish between a random process and randomness in a sequence of outcomes generated by a random process (Zabell, 1992). Previous research into perceptions of randomness has not always been clear about this distinction (Nickerson, 2002), but some writers have discussed the issue explicitly (Falk and Konold, 1997; Wagenaar, 1991). The position that randomness is a property of a process rather than of the outcomes was adopted by Wagenaar (1991), and related to this is the view that any outcome from a random process is considered a random outcome (Pollatsek and Konold, 1991). However, it is only by observing outcomes from a process that one can judge whether the process is random.

Johnston-Wilder (2006a) suggested that randomness needs to be seen as 'dynamic', since a permanent printout of a random sequence loses the essence of what it is to be random. He reported his own struggle to see random number tables as random since "whenever I opened the book the sequence was the same". In the present research we see randomness as a model describing a process and providing an explanation for observed outcomes. A sequence of observed outcomes is described as 'random' if it can be considered to have arisen from a process modelled by 'randomness'.

Stimulus tasks commonly used in previous research may be classified into two categories (Falk and Konold, 1997). In a generation task, subjects make up a random sequence of outcomes to simulate outcomes from a random process such as 'tossing a coin'. In a recognition task, subjects might select the 'most random' of several sets of results. Falk and Konold suggest (1997) that recognition tasks "may be more appropriate for revealing subjective concepts of randomness" because "a person could perceive randomness 'accurately' and still be unable to reproduce it" (p. 302). Indeed there is clear suggestion from many studies (Nickerson, 2002; Shaughnessy, 1992) that people are generally not good at generation tasks, typically producing fewer long runs, and more alternations between outcomes, than would be expected from a random process (Falk and Konold, 1997). Other studies have used recognition tasks to explore what sequences people consider to be maximally random; these again show that people tend to identify randomness with sequences having an excess of alternations between outcomes (Falk and Konold, 1997).

LOCAL AND GLOBAL PERSPECTIVES ON RANDOMNESS

Pratt studied ways in which children aged 10 and 11 years articulated their ideas and beliefs as they worked in a carefully designed computer-based domain (Pratt 1998, 2000). He distinguished two categories of meaning for randomness expressed by children: 'local meanings' were related to uncertain behaviour of the process and focused on "trial by trial variation", while 'global meanings' evolved as children recognised the importance of observing a larger number of trials and discerned long run features of distribution. The terms 'local' and 'global' also appear in Ben-Zvi and Arcavi's account (2001) of 'local and global views of data', where local understanding typically relates to a few values, while global understanding relates to general patterns in the dataset.

When analyzing data from the early stages of students' work, Pratt noted that students did not articulate global meanings, but described situations entirely in terms of local meanings. Only later, after working with a specially designed computer environment, did the children begin to express meanings for longer term aggregation. As the children worked in this environment, mending broken simulations of random generators such as coins and dice, their attention was drawn to the different behaviour of the generators in the short and long term. Pratt saw the transition from local to global meanings and analysed in some detail the complexity of that process. He did not, however, report movement from global to local meanings. Perhaps children of this age were not able to articulate global meanings until their thinking had been perturbed by the use of the computer-based tools. Alternatively, perhaps it is implicit in Pratt's tasks and probes that no such switch was either anticipated or looked for. Similarly, Ben-Zvi and Arcavi (2001), in a study of pupils' developing understanding of data and data representations, saw the transition to learning to look for global understanding of data as part of a developmental process.

The focus of the current paper is on whether we can identify transition in the opposite direction, from global to local, and to explore the nature of such movement.

METHOD

We anticipated that we might find evidence of such a shift from global to local by studying older students. We recognise that such an approach would not clarify whether the lack of evidence in Pratt's study for the articulation of global meanings at age 11 was as a result of the method he used or the naivety of the learners. Nevertheless, it seemed important to establish whether or not shifts from global to local could happen before we considered what might stimulate such shifts.

Eighteen learners, aged from 13 to 17 years, undertook clinical interviews for about an hour each. Interviewees were pupils at local secondary schools (13-18), selected by an experienced mathematics teacher in the school as pupils who would be able to express their ideas confidently. They had agreed to participate in the study, and their parents had also given permission.

Interviewees worked on tasks using three unusual dice: biased, spherical and cracked. The biased die looks like a standard cube, except it has two faces labelled 5 and no face labelled 3. It has a weight in the face labelled 1, biasing it towards showing 6. The spherical die is hollow and marked symmetrically with numbers 1 to 6. It contains a small bead so that, when rolled on a flat surface, it stops with one of the six numbers uppermost. If it is correctly balanced, each of the six outcomes is equally likely. The cracked cubical die has a split running across the face labelled 6, and spreading partway across the faces labelled 2 and 5. Since interviewees used this die after their experiment with the biased die, it was expected that they would consider that it might also be biased.

In each task, the learner was first asked to comment on the appearance of the die and to consider how it might behave when rolled several times. The learner was then invited to roll the die a few times before commenting on the observed outcomes. Learners were encouraged to discuss their ideas throughout and we watched their behaviour closely. If a learner appeared to show concern about a run of outcomes, or even an individual outcome, they were invited to explain what they were thinking. It was hoped that using three different dice would increase the learner's awareness of what they expected from each die, and their willingness to articulate their assumptions. In particular, it was hoped that the tasks would provoke learners to talk about how to recognise equally likely outcomes and whether these were necessary for the die to be considered to behave 'randomly'.

Each interview was audio-taped and transcribed in detail for analysis. A commentary was written around extracts from the transcript to provide a detailed account of key moments. Finally, common themes were identified across the interviews. At this stage of the analysis, the relationship began to emerge between two distinct ways of thinking about random outcomes: local and global perspectives.

THE DATA

Analysis of data from the present study has identified two ways of thinking about randomness, which are clearly related to Pratt's local and global meanings. Learners were seen to shift between these two perspectives, from local to global and back to local. Sometimes these shifts appear frequent and rapid. These differing perspectives reflect the learner's focus of attention.

In this paper we report data from two cases, Ben and David, selected to illustrate issues surrounding relationships between local and global perspectives. Interviewees' shifting attention is not apparent in isolated incidents but rather needs to be tracked through a sequence of interactions over a period of time. This requires discussion of extensive passages from the commentaries, and there is not space to discuss more than two cases. The quoted excerpts show the actual words spoken by the learners, even when these are ungrammatical.

Ben's story

In these excerpts, Ben (age 15.7) was using the spherical die. In the first seven throws, he observed {5 5 1 4 1 6 1} and he looked for patterns in the sequence.

1. Ben: …we haven't had any 3s or 2s, so it could be one of those, but – well, it'll probably be another number than a 1.
2. Interviewer: Why?
3. Ben: Just from following the pattern. If it wasn't a die, that's what I'd say.

Ben noted that any outcome was possible. He went on to seek an explanation for the absence of 2s and 3s, looking at the generating process from a local perspective.

4. Ben: It might be the way I'm throwing it though. Or when I picked it up, I'm throwing it the same way. Or it could just be chance.

Out of concern about lack of 2s and 3s, Ben checked the labelling of the die. When the fourteenth outcome was 3, he cheered! He now attended to physical factors affecting the outcomes, deliberately playing with the die between rolls, and feeling the weight moving inside the sphere. When asked how he would know this was a fair die, he expressed a global perspective based on his prior belief about a fair die.

5. Ben: …You just have to keep rolling it. It should in the end even out if it's a fair dice. If it's not a fair dice it'll… keep on staying away from the 2s and 3s, like it is at the moment.
6. Interviewer: Are you worried about it being fair?
7. Ben: …No, not really. …It could just be chance. If there's a 1 in 6 chance of getting each different number... I just haven't got a 2 yet, which is strange. Although I'll probably get a 2 now, if I roll it...

Ben switched rapidly between contrasting views about this die. Although he understood the need for more trials (characteristic of a global empirical perspective), and expressed a distributional belief that the die would be fair, he still used a local perspective in looking for the first occurrence of a 2 and was concerned when he had not seen it in fifteen throws. Globally, he looked for a frequency distribution to match his prior belief, and accepted that "It could just be chance". By changing the focus of his awareness, Ben arrived at contrasting explanations for the absence of 2. After 17 throws without a 2 {5 5 1 4 1 6 1 1 5 4 4 5 1 3 6 5 4 5 5}, Ben was quiet, and experimented with the die, rolling it in his hand without talking for 13 seconds, before commenting.

8. Ben: It seems pretty fair. But it depends what happens when you roll it.

He seemed to experience a tension between apparent 'fairness' of the process, and imbalance in the outcomes. Then he rolled a 6, but he wanted the die to show a 2. It was as though he wanted to remove the anticipation of waiting for a 2 to occur, and by experimenting with the way he rolled the die he was trying to make it happen.

9. Ben: …oh land on a 2.

On the next throw Ben rolled the die, and got a 2! He was excited because his experimenting with how to roll the die had coincided with rolling a 2, and so he began to think he could control the outcomes. Then he went quiet again, until he was asked what he was thinking about. He commented that the axis of rolling the die did not explain the outcomes as 2 and 3 were not opposite to each other on the die.

10. Ben: Just seeing… if I was always rolling it in a way so it only lands on 6, 5 1, 4. But that wouldn't work, or make sense… it stays away from 2s and 3s, but it won't cos they're not next to each other – but they are.

Ben was again reasoning in a local perspective, trying to explain the short sequence of outcomes observed. But in the next sentence he began to search for a causal explanation that may have impact on the chance of certain numbers appearing.

11. Ben: It might be weighted more heavily on the 2 and the 3, on the inside, I was just thinking. If the weight's heavier there it will be less likely to turn that way.

Ben rolled a 1 and remarked that the die seemed more random now.

12. Ben: The more you do it, you know, the more different… But at the start it was all the same. So the more you do it the better the results you get, I suppose.

Perhaps in line 11, consideration of probabilities had signalled a possible shift to a global perspective. By line 12, it seems Ben had indeed found a global explanation and he tried to stabilise this idea in his mind. The next throw produced a 5.

13. Ben: …It should, unless it's weighted, be completely random. But at the start it just seemed to be 5s and 1s. But then it… just got a lot more mixed as it went down, so I suppose… it's just… more and more of a dice and, sort of less chance that the odd number will count for so much. You got a couple of 5s at the beginning, then, later as you go on, you'll get more of the other numbers as well. In theory, I think. …Although I haven't got that many 2s still.

As he moved towards a stable global perspective, Ben was holding in tension the two contrasting ideas of *randomness* (by which he meant equiprobability) and *bias* – and he expressed them alternately. These were the apparently conflicting global interpretations for Ben: prior belief and a global frequentist view, possibly emerging from the aggregation of the observed outcomes. As soon as he expressed the idea of randomness, he reverted to discussing the bias. He rolled another 5 and reverted to a local perspective.

14. Ben: But I have got quite few 5s I think. …But that could just be the way I'm rolling it.

Over the next few throws, Ben's concern about bias diminished as he obtained more 2s and 3s. He remarked again on the apparent randomness.

15. Ben: Maybe it's just… I suppose it could just be a completely fair dice… It does have quite a few 5s, but that just might be me rolling it, rather than the dice would be weighted or something.

David's story

In the following account of David (age 14.1) working with the biased die, we show how David's attention moved towards reconciling two global perspectives: his prior beliefs about the probability distribution of the six possible outcomes and the emerging sequence of observed outcomes.

David observed two 6s and a 4 in the first three throws, but stated that the outcomes were equally likely.

16. David: It's all the same probability, but it gave me more 6's than any other number.

After two further 6s, David was concerned that the die "always seemed to land on 6". His attention was shifting from his global belief about the probability distribution to the short sequence of observed outcomes – a local perspective. When the seventh throw gave a 5, David tried to articulate a pattern that he thought was emerging.

17. David: It seems to land on higher numbers than lower numbers.

Although his attention was on the short sequence of observed outcomes, he tried to express a new global view from the pattern he saw {6 4 6 6 6 6 5}. However, he still had the idea that the outcomes should be equiprobable. After two further 6s he suggested that 'chance' would correct the imbalance by producing a lower outcome.

18. David: I think it'll be a lower number next… because there's been too many higher numbers. It could be any of the six numbers. I think it might be a lower one.

David was persisting with the idea that the outcomes were equally likely, but he tried to modify the behaviour of a 'chance' process to be self-correcting. After rolling another 6, David took this idea to an extreme, choosing the lowest value available as his prediction to maximise the degree of correction.

19. David: (Silence for 10 seconds) I think it might be a 1… because it's landed on 4, 5 and 6 and on each of the sides it's close to 2, 4 and 1.

The long pause suggests that David was unsure how to respond. He had observed ten outcomes {6 4 6 6 6 6 5 6 6 6}. When the eleventh throw gave a 5, David changed his position, using recent outcomes as a guide to what might happen next.

20. David: (Throws) 5… I think it might land on another 6. (Laughs)

21. Interviewer: Why have you changed your mind?

22. David: Because… it seems to always land on high numbers, and I'm not sure why, it just always seems to land on a high number. It hasn't landed on any under 4 has it?

From here onwards, the emerging global distribution was so different from the idea of equiprobable outcomes that David appeared to accept the die was biased. However, even after two further 6s, he still struggled to express this new view. When he was asked if he still thought that the next outcome could be any number, he restated, hesitantly, the idea that the outcomes were equally likely.

23. Interviewer: You started off by saying that it could be any number… Do you still think it could be any number?

24. David: I think it can be any number yeah…, cos there's a one in six chance of getting every number there.

At this point, David again examined the die closely, trying to explain the emerging global distribution. For the first time, he spotted that the die was incorrectly labelled – an extra 5 in place of the 3 – but this did not explain his observations.

25. David: Yeah but that means it should land on 5 more, but it doesn't. (Silence 9 seconds) But there is still one in six chance of getting a 6.

Again, his argument reverted to equiprobability, although he seemed to be actively seeking a reason to suggest that the die was biased. The disparity between the observed outcomes at the local level and his prior belief in equiprobability was now driving his search for an explanation, but he had not yet abandoned his prior belief. When he rolled another 6, David picked up the die to examine it again. When asked how many 6s he had observed, he counted thirteen 6s in sixteen throws. Even now, he clung to the idea that the faces should be equally likely and, at the same time, that chance would correct the imbalance of outcomes.

26. David: I'm not sure… There should be a one in six chance of getting a 6… I'm hoping it's going to land on a low number. (Laughs)

After another 6, he finally expressed a global view that the die was biased.

27. David: I think the probability of getting 6 is higher now. Because just of the outcome. I'm not sure why.

Although David was convinced that the die was biased, he had no explanation for his global view. From now on, he examined the die after each throw until he spotted the weight. Then he quickly explained the observed distribution of outcomes.

28. David: (Examines the die again) Is it cos that bit there – is metal? So, it's going to… put more outcome onto 6, put more chance onto it… Well, I'm not sure cos… there's two 5s, but it would have to… go on the other side of that… metal part. And that only happens when it slides across. But every time I actually roll it, it always lands on a 6. So I think it might be that, it's heavier, so it's landing down further, and there's more force going down, so it's going to stick down on it.

At last, David was able to construct, in his own terms, a global view of the 'probability' distribution that he could reconcile with what he observed at the local level. Because he had observed 6 so often, he argued that the weight "is heavier" and the die is "going to stick down on it". Until he found this explanation, he could not be comfortable about rejecting his prior belief, and he could not reconcile the observed outcomes with his global sense of distribution.

David's response to the biased die was unusually protracted. He appeared to have been deeply committed to his prior idea that the die should be fair and he continued to seek justification for this view, even when he had observed thirteen 6s in sixteen throws. His attempts to justify that his observations could arise from a fair die indicated that, although he was aware that outcomes from a random process might not be representative of the long run, he was possibly unaware of how much variability to expect in the outcomes of a fair die.

There were further examples of David shifting between the local and the global perspectives when he was working with the other activities. However, when he worked with the spherical die and the cracked die, his prior beliefs were supported by the emerging empirical distribution, since the observed frequencies were not very different from his prior belief that the outcomes would be equally likely.

CONCLUSIONS

Analysis of transcripts in this study has showed that learners think about random outcomes using two distinct and contrasting perspectives: local and global, which reflect different focuses of attention. In the local perspective, attention is on the uncertainty of the next outcome and ephemeral patterns that appear in short sequences of outcomes. The learner does not aggregate outcomes or think in terms of a distribution. In the global perspective, the learner is aware of a distribution of outcomes, either empirically, as an emerging frequency distribution of observed outcomes, or in terms of prior beliefs about the generating process (for example, when rolling a die, expecting the outcomes to be equally likely). These two perspectives are clearly related both to Pratt's local and global meanings of randomness (Pratt, 1998) and to Ben-Zvi and Arcavi's (2001) local and global understandings of data. However, evidence in the present study had shown that learners' attention shifts frequently, and sometimes rapidly, between these perspectives, from local to global and back to local.

Ben's ideas were strongly affected by short run behaviour of the spherical die. When the sequence of recent outcomes did not include one or two of the possible outcomes, he tried to explain the apparent bias. When the missing outcomes had appeared once or twice, he described the behaviour of the die as "random". Sometimes "random" was "the absence of pattern", and this cue was switched on and off by short-term changes in the sequence of outcomes. Ben's interpretation of the outcomes was also influenced by the fact that he did not know how much variability to expect from a fair die. For example, he did not know how many throws he might need to wait until all six outcomes had appeared at least once, or how often should the most commonly occurring outcome appear in the first *n* throws of the die. Therefore Ben could not judge whether he had seen too many 5s, or whether the waiting times he observed before the first 3 and the first 2, were appropriate in a fair die. To refine his judgement of whether a die was fair, Ben needed intuitions about variability. Understanding of variability is important in reconciling the local and the global views of randomness. In all the interviews, shifting attention between local and global perspectives in both directions was common, and the phenomenon seemed to be fuelled by a desire to draw conclusions from short sequences of outcomes. This in turn seems to be related to a poor understanding of variability.

The interviews also provided insight into the process by which an individual begins to reconcile what is seen in the local perspective with an emerging global view. With the spherical die (Ben's story, for example), the results of throwing the die do not seem to demand any causal explanation. Even so, Ben's articulation in line 11 "the more you do it the better the results you get" seems to have a causal root to it.

Where there was mismatch between a strongly held prior belief and the long run frequency distribution, the tension within the individual was seen to mount with each successive outcome. David's work with the biased die stands out in this regard. His attention shifted rapidly between the conflicting perspectives, and he was uncertain about what to attend to. In the case of the biased die (David's story), the strange data easily provokes causal explanations. Perhaps surprisingly, those causal explanations seem to stimulate a global perspective.

However, is it so surprising? Pratt's study (1998, 2000) showed that global meanings constructed when working within his computer-based microworld had a causal nature and recent work (Prodromou & Pratt, in press) has sought to marshal such causal meanings in making sense of distribution. Indeed, Piaget argued in his seminal work (Piaget and Inhelder, 1975) that the organism eventually invents probability in order to operationalise randomness. When we take a global perspective, we recognise long term predictability as opposed to short term unpredictability. This predictability can be understood in terms of causal forces such as the weight in the die or the effect of many throws or a mixture of the two.

We have provided evidence that these students, aged 13 years and above, shifted between local and global perspectives in both directions, adding to the findings by Pratt of shifts from local to global. Indeed, such shifts were prolific. It remains to be seen whether such evidence can be found for younger students. This study also suggests that causal factors may play an important part in such shifts.

NOTES

1. This paper refers to data and analysis from the first author's doctoral thesis (Johnston-Wilder, 2006a) and is partly based on an earlier paper (Johnston-Wilder, 2006b).

REFERENCES

Ben-Zvi, D. and Arcavi, A. (2001). "Junior high school students' construction of global views of data and data representations". *Educational Studies in Mathematics*, 45, 35-65.

Falk, R. and C. Konold (1997). "Making Sense of Randomness: Implicit Encoding as a Basis for Judgement". *Psychological Review* 104(2), 301-318.

Johnston-Wilder, P. (2006a). "Learners' Shifting perceptions of Randomness", Unpublished PhD Thesis, Centre for Mathematics Education, Open University, January 2006.

Johnston-Wilder, P. (2006b). "Learners' shifting perceptions of randomness", in Hewitt, D. (ed), *Proceedings of the British Society for Research into Learning Mathematics*, vol. 26, No2. ISSN 1463-6840

Nickerson, R. S., (2002). "The production and perception of randomness". *Psychological Review*, 109(2), 330-357.

Piaget, J. and Inhelder, B. (1975). *The Origin of the Idea of Chance in Children*, London: Routledge and Kegan Paul

Pollatsek, A. & Konold, C. (1991). "Randomness is well enough understood to be misunderstood". *Journal of Behavioral Decision Making*, 4(3), 218-220.

Pratt, D. (1998). "The Construction of Meanings In and For a Stochastic Domain of Abstraction". Unpublished PhD Thesis. Institute of Education, University of London, May 1998. Available as pdf from http://fcis1.wie.warwick.ac.uk/~dave_pratt/page11.html

Pratt, D. (2000). "Making Sense of the Total of Two Dice", *Journal for Research in Mathematics Education*, 31(5), 602-626.

Prodromou, T. & Pratt, D. (in press). "The Role of Causality in the Coordination of Two Perspectives on Distribution within a Virtual Simulation", *Statistics Education Research Journal*.

Shaughnessy, J.M. (1992). "Research in probability and statistics: Reflections and directions". In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning*. New York: NCTM & MacMillan, 465-494.

Wagenaar, W.A. (1991). "Randomness and randomizers: Maybe the problem is not so big. Commentaries on 'Psychological conceptions of randomness'". *Journal of Behavioral Decision Making*, 4(3), 220-222.

Zabell, L. S., (1992). "The quest for randomness and its statistical applications", in F. Gordon and S. Gordon, (Eds.), *Statistics for the twenty-first century*, Washington DC: Mathematical Association of America, 139-150.

TRANSPARENT URNS AND COLORED TINKER-CUBES FOR NATURAL STOCHASTICS IN PRIMARY SCHOOL

Laura Martignon, Kathryn Laskey and Elke Kurz-Milcke

Institute for Mathematics, Pädagogische Hochschule Ludwigsburg

and the Krasnow Institute, George Mason University

We join the camp of mathematics educators who claim that children should receive early training in stochastic thinking, although their training – at such stage – may only be based on a heuristic understanding of stochastic phenomena. The term "heuristic" is to be taken in the sense propagated by scientists like Einstein and Polya, which is not as "a rule of thumb" but as a correct yet partial approximation of the normative approach. We present a review of empirical results supporting our claim and propose to guide children to construct stochastic situations enactively.

EFFECT OF REPRESENTATION FORMATS IN PROBABILISTIC REASONING TASKS

The theoretical framework of this paper is that provided by Hasher & Zack's work on humans' automatic recording of frequencies in the environment combined with Barsalou's mental concept simulators (1999) and the theory of heuristics for inference proposed by Gigerenzer, Todd and the ABC Group (1999). Barsalou's simulators implement a basic conceptual system that represents types, supports categorization, and produces categorical inferences. Productivity results from integrating simulators combinatorially and recursively to produce complex simulations. In this framework mathematics education of young students consists in the education of an inner representation space where mental simulators of mathematical processes are implemented (e.g., "imagine drawing from an urn"). Mathematical intuitions in general and probabilistic intuitions in particular are thus conceptually replaced by heuristics for inference combined with mental simulations that are part of an adaptive toolbox (Gigerenzer et al. 1999). For such mental simulators *natural* frequencies (see below for a characterization), are more *ecologically rational* than percentages and probabilities (1995). We are beginning to find confirmations from cognitive neuroscience of this thesis. In fact, we have collected evidence suggesting that the regions of the brain which are active when we perform probabilistic inferences by means of natural frequencies differ from those that are active when we solve probabilistic problems with percentages or probabilities. This is true even for expressions such "1 out of 4" as compared with expressions such as "0.25" or "1/4".

Initial experimental results substantiating the above hypotheses support an emphasis in schools on *natural* representation formats for probabilistic information. Here, the term "natural" means arising either directly from enactively constructing subcategories of a population by partitioning it sequentially in nested subsets and determining the proportions of the subcategories thus formed, or mentally simulating these same processes. Our emphasis is not meant to replace instruction in percentages or measure theoretic probability. On the contrary, we view early school interventions as a means to prepare young children for later instruction in working in the formal mathematics of probability. Enactive learning approaches, we claim, can teach young children to reason with proportions of counted items, thus making use of natural representation formats to develop intuitions. Probabilistic inference in secondary school can then make use of this previously acquired substrate by anchoring probabilistic reasoning in “translations” of probabilities into natural frequencies. The advantages of such “translations” have been empirically tested in interventions in secondary school (DFG Project in BIQUA, Ma-1544/1-4 and Bi-384/4-3).

Cognitive Processes in Probabilistic Reasoning Tasks

As awareness grows of the importance of uncertainty in everyday life and in public affairs, concern also grows about the competence of the citizenry to process uncertainty in a sound and effective manner. An alarmingly large proportion of the public cannot make effective use of probabilistic information. The German newspaper “Süddeutsche” (Süddeutsche Zeitung Magazin, 31.12.1998) asked 1000 Germans, what they think is the meaning of 40%: a quarter, 4 out of 10, or every 40^{th}. Only 54% knew the correct answer, which is “4 out of 10”. A burgeoning literature has documented disparities between the results of unaided judgment and the prescriptions of the probability calculus (KAHNEMAN, SLOVIC AND TVERSKY, 1982). Summarizing the literature on human performance on probabilistic reasoning tasks, Gould (1992) commented: “*Tversky and Kahneman argue, correctly, I think, that our minds are not built (for whatever reason) to work by the rules of probability.”* Yet, recent re-examinations of the literature on human performance on tasks involving uncertainty have concluded that to a large extent, the negative results can be explained by discrepancies between the environment and tasks on which present-day humans perform so poorly, and those faced by our ancient forebears (e.g., GIGERENZER, et al., 1999). The question is then: *Would it be possible to improve the public's skill at probabilistic reasoning by matching pedagogical strategies adaptively to cognitive processes during early phases of education, thus providing anchoring mechanisms and "translation" heuristics for the phase when more formal representations are taught?* We build on a base of existing results on the cognitive mechanisms underlying probabilistic reasoning. Our research was originally motivated by an important type of probabilistic reasoning task known as “Bayesian reasoning.” A prototypical Bayesian reasoning task involves using evidence about an uncertain proposition to revise our assessment of the likelihood of a related proposition. The following example is drawn from a recent article by Zhu and Gigerenzer (2006) that examined children’s ability to perform Bayesian reasoning. The context was a small village in which “red nose” was a “symptom” of “telling lies”. The task required the children to relate the proposition “having a red nose” to the proposition “telling lies.” Specifically, they were given information about the probability of “red nose” conditioned on “liar” and that of “red nose” conditioned on "non-liar," as well as the incidence of liars in the village. They were then asked to establish the chance that someone with a red nose tells lies. The performance of children (fourth graders) improved substantially when the probabilistic setting was replaced by a setting in which the cover story reported the natural frequencies involved, i.e., in terms of a sequential partitioning of nested sets and their proportions. Humans, children and adults, appear to be adapted to a "natural" sequential partitioning for categorization. The term "natural" means that these proportions (i.e., relative frequencies) are perceived as being obtained by the mental simulation of counting. ATMACA AND MARTIGNON (2004) conjectured that different neural circuits are involved in the natural frequency and probability versions of the Bayesian task. They reported experimental results that support their conjecture. Subjects were given tasks by slide projector, and solved them mentally with no writing allowed. Information was collected on correctness of solutions and the time to solution. The experiment made use of a response mode called *result verification* or *result disparity* (KIEFER and DEHAENE, 1997): Subjects are presented with a proposed solution and asked to judge as quickly as possible whether it is correct or incorrect. ATMACA AND MARTIGNON found that subjects needed significantly longer times and produced significantly fewer correct answers, for the tasks given in probability format versus those given in the natural frequency format. In one experiment 110 participants were exposed to typical Bayesian tasks with "three branches" and tasks with "four branches" as represented below:

Three branches: 10 out of 1000 children have German measles . Out of the 10 children who have German measles, all 10 have a red rash. *Of the 990 children without German measles, 9 also have a red rash. How many of the children with a red rash have the German measles?*

Four branches: *10 out of 1000 car drivers meet with an accident at night. Out of the 10 car drivers who meet with an accident at night, 8 are intoxicated. Out of the 990 car drivers who do not meet with an accident at night, 40 also are intoxicated. How many of the car drivers who are intoxicated actually meet with an accident at night?*

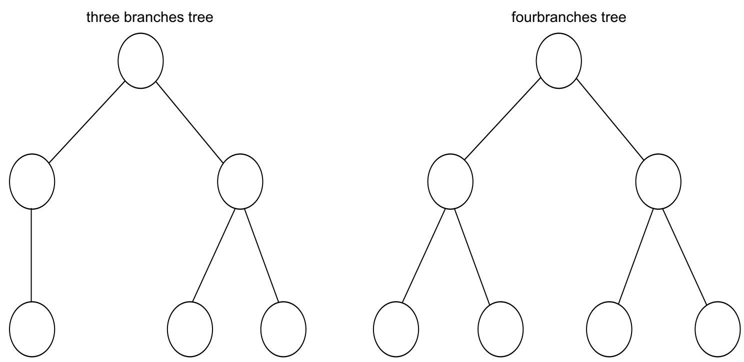


Figure 1: Bayesian Task

Results of the experiment are summarized in the following Figure:

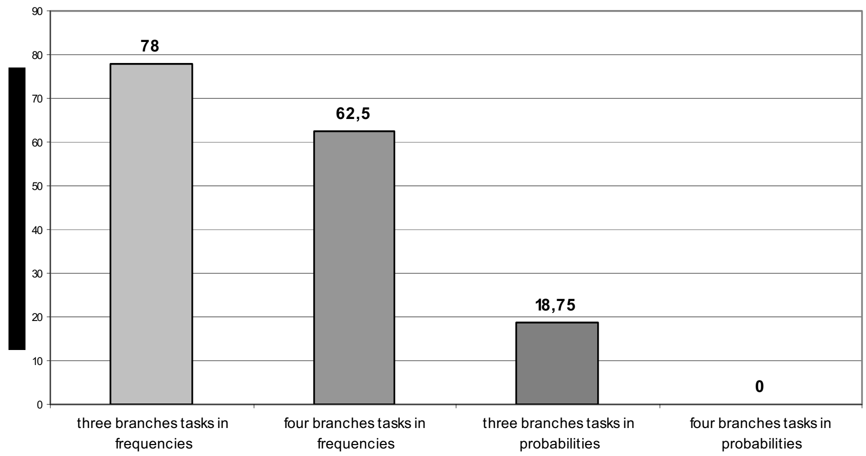


Figure 2: Percentage of Subects Solving Task Correctly

Perception of frequencies of occurrences, this result suggests, could be a mechanism, or at least part of a more complex mechanism, that enables fast and effective decisions in uncertain situations, because "*...natural selection (...) gives rise to practical cognitive mechanisms that can solve (...) real world problems...*" (FIDDICK & BARRETT, 2001, S. 4). DEHAENE (1997) wrote in a similar context: "*Evolution has been able to conceive such complex strategies for food gathering, storing, and predation, that it should not be astonishing that an operation as simple as the comparison of two quantities is available to so many species).*" (DEHAENE 1997, p. 27)

The Automatic Processing of Frequencies

The automatic perception of frequencies of occurrences was described by HASHER AND ZACKS in the 70s: "Operations that drain minimal energy from our limited-capacity attentional mechanism are called automatic; their occurrence does not interfere with other ongoing cognitive activity. They occur without intention and do not benefit from practice". Certain automatic processes, we propose, are ones for which humans are genetically "prepared". These processes encode the fundamental aspects of the flow of information, namely, spatial, temporal, and frequency-of-occurrence information." (HASHER AND ZACKS, 1979, p. 356). Humans are known to be sensitive to frequencies even when they do not pay attention to them. In experiments, participants performed well when remembering frequencies of events, even when they had no reason to expect a memory test at all (ZACKS, HASHER AND SANFT, 1982), and there seems to be evidence that they did not count the events (COREN AND PORAC, 1977). As reported by HASHER AND ZACKS (1984), several experiments have shown that neither activated intention, nor training, nor feedback, nor individual differences such as intelligence, knowledge or motivation, nor age, nor reductions in cognitive capacity such as depression or multiple task demand, have an influence on the processing of frequencies of occurrence information. That is to say, there are strong hints that the human brain adapted to the processing of frequencies during evolution. Furthermore, several experiments have provided evidence that "(…) various animal species including rats, pigeons, raccoons, dolphins, parrots, monkeys and chimpanzees can discriminate the numerosity of various sets, including visual objects presented simultaneously or sequentially and auditory sequences of sounds" (DEHAENE ET AL., 1998, p. 357). According to models of animal counting presented by MECK AND CHURCH (1983), numbers are represented internally by the continuous states of an analogue accumulator. For each counted item, a more-or-less fixed quantity is added to the accumulator. The final state of the accumulator therefore correlates well with numerosity, although it may not be a completely precise representation of it. (DEHAENE, 1992) This model explains the observation that animals are very good in handling small quantities, while performance degrades with the increase of magnitude. During this counting process, "the current content of the accumulator is used as a representative of the numerosity of the set so far counted in the decision processes that involve comparing a current count to a remembered count." (GALLISTEL AND GELMAN, 1992, p. 52) That is to say, the current content of the accumulator represents the magnitude of the current experienced numerosity, whereas previously read out magnitudes are represented in long-term memory. This enables comparison of two number quantities, which is essential for frequency processing. The comparison of number processing abilities in animals and human infants leads to the conclusion "that animal number processing reflects the operation of a dedicated, biologically determined neural system that humans also share and which is fundamental to the uniquely human ability to develop higher-level arithmetic." (DEHAENE ET AL., 1998, p. 358). Several studies report abilities of frequency perception in kindergartners and elementary school children in the range of grades 1 to 6 (HASHER AND ZACKS, 1979; HASHER AND CHROMIAK, 1977). And the findings of abilities in numerosity discrimination in infants and even newborns (ANTELL AND KEATING, 1983) "may indicate that some capacity for encoding frequency is present from birth." (HASHER AND ZACKS, 1984, p. 1378).

From Absolute Numerical Quantities to Natural Frequencies

Animals and humans could not survive if they had only developed a sense for absolute frequencies without a sense for *proportions of numerical quantities* for inference. "Are all red mushrooms poisonous, or only some of them? How valid is red colour as an indicator of poison danger in the case of mushrooms?" Whereas non-precise estimates may have been sufficient for survival in ancient rural societies, answering this type of question by means of well calibrated inferences is vital in modern human communities. Successful citizenry requires this type of competency and, it is our conviction, elementary school should provide tools for successful quantified inferences. In order to establish whether one cue is a better predictor than another (e.g., whether red colour is a better predictor than white dots for poisonous mushrooms) we need a well tuned mental mechanism that compares proportions. Little is known so far, in the realm of cognitive neuroscience, as to which brain processes are involved in proportion estimation and proportion comparison. The more rudimentary instruments for approximate inference that humans share with animals must be based on some sort of non-precise proportion comparison (GIGERENZER, ET AL., 1999). But when do infants in modern societies begin to quantify their categorization and when can they be trained in quantified proportional thinking? Although fractions are the mathematical tool for describing proportions, we envisage an early preparation of children's use of numerical proportions without "normalizations" before they are confronted with fractions. The results by Piaget and Inhelder on the understanding of such proportions in children motivated two generations of researchers in developmental and in pedagogical psychology related to mathematics education. We cite here one direction in particular, which has been fundamental to our work (Koerber, 2003). In a series of well designed experiments, STERN, KOERBER and colleagues (STERN ET AL. 2002) demonstrated that third-graders can *learn to abandon* the so-called *additive misconception*, in which children respond with "9" instead of "12" to "3 : 6 = 6 : ?". In these experiments, children were asked to compare mixtures of lemon and orange with respect to their intensity of taste. The training involved using a balance beam or graphs to represent juice mixtures, and moving the pivot to represent changes in proportions. At the end of a short training (2 days at most) children showed improvement in proportional thinking, The results of STERN and her school thus provide evidence of third-graders' aptitude to learn proportional thinking when provided with adequate instruction.

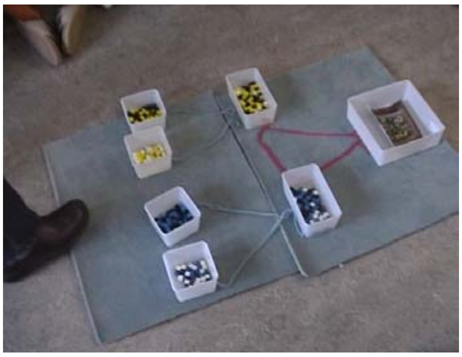
Cognitively Natural Representations and Task Performance

Stern's results can be combined with results of Gigerenzer and his school at the interface between pedagogy and cognitive science in the search for pedagogical approaches that tap into cognitively natural representations. The *natural frequency* representation for Bayesian reasoning tasks is based on information that can be gained by "naturally"counting events in an environment, and therefore taps into very basic human information processing capacities. "*Natural sampling is the way humans have encountered statistical information during most of their history. Collecting data in this way results in natural frequencies*." (HOFFRAGE, GIGERENZER, KRAUSS & MARTIGNON, 2004) The term "natural", as has been pointed out, signifies that these frequencies have not been normalized with respect to base rates. Probabilities and percentages can be derived from natural frequencies by normalizing natural frequencies into the interval [0,1] or [0, 100], respectively; however, this transformed representation results in loss of information about base rates. Consider the following examples from HOFFRAGE, GIGERENZER, KRAUSS & MARTIGNON, 2004:

> Natural frequencies: *Out of each 100 patients, 4 are infected. Out of 4 infected patients, 3 will test positive. Out of 96 uninfected patients, 12 will also test positive.*

> Normalized frequencies: *Out of each 100 patients, 4 are infected. Out of 100 infected patients, 75 will test positive. Out of 100 uninfected patients, 12.5 will also test positive.*

Because normalized frequencies filter out base rate information, they make the Bayesian task of inferring a posterior probability from evidence more difficult. The inference that most positives are false positives can be read directly from the natural frequency representation, while it must be obtained via a non-trivial calculation from the normalized frequency representation. Training young children enactively with natural proportions, we claim, enables them to make use of simple heuristics for dealing with probabilities. We have worked with fourth graders, preparing them to solve one of the mathematical test items of PISA 2003 namely: *Consider two boxes A and B. Box A contains three marbles, of which one is white and two are black. Box B contains 7 marbles, of which two are white and five are black. You have to draw a marble form one of the boxes with your eyes covered. From which box should you draw if you want a white marble?* Only 27% of the German school students were able to justify that one should choose Box A. Mathematically correct statements regarding why and when larger proportions in samples correspond to larger chances in the populations require serious amounts of conceptual work. In early grades, the consensus is that one should focus on intuition and competency rather than on formal mathematics. In other words, we should provide students with: (1) Basic stochastic modelling skills with natural representation formats and (2) Simple heuristics for operating with these formats. In this spirit, MARTIGNON AND KURZ-MILCKE (2005) and KURZ-MILKE AND MARTIGNON (2005) have designed a program that develops and encourages the natural frequency representation through the use of enactive learning. In playful yet structured activities, children use coloured plastic cubes called tinker-cubes to represent individuals that make up a population. Different colours represent different attributes (e.g., red cubes for girls; blue for boys). The cubes can be attached one to another, allowing representation and multi-attribute encoding (e.g., a red cube attached to a yellow cube for a girl with glasses; a blue cube attached to a green cube for a boy without glasses, a red cube attached to a green cube for a girl without glasses, and so on). Children collect the tinker-cubes into plastic urns that represent populations. In this way, they gain concrete visual and tactile experience with individuals with multiple attribute combinations and how they can be grouped into categories and subcategories. Recent exploratory studies of fourth-grade children indicate that children are both enthusiastic and successful when constructing these representations of categories and sub-categories in nested sets, especially when the populations are personally meaningful (e.g., "our class"). They can easily "construct" answers to questions like "how many of the children wearing glasses are boys?"



2. Constructing our class
3. Enactive Bayesian Reasoning
4th – graders
(Kurz-Milcke&Martignon, 2005)

Figure 2

Figure 3

In another activity, children enact a model of proportional reasoning by constructing so called similar urns to represent equivalent proportions (e.g., an urn containing 2 red and 5 blue tinker-cubes – denoted by U (2:5) - is similar to an urn containing 4 red and 10 blue tinker-cubes). By this activity the children learn basic urn arithmetic. For instance, fourth graders in three classes of a school in Stuttgart successfully learned to solve the two boxes task described above, where U(1:2) is compared with U(2:5) by first constructing an urn U(2:4) similar to U(1:2) and then easily comparing U(1:2) and U(2:5). These tasks contain first elements of elementary probabilistic reasoning in general but also of Bayesian reasoning at a heuristic level. They represent a preparation for understanding both of fractions and − at a later stage − of probabilities. Empirical longitudinal studies have now been designed to confirm the hypothesis that mastery of these tasks in the younger grades should support better performance on stochastics questions in the later grades.

REFERENCES

Atmaca, S. & Martignon, L. (2004) Hat sich unser Gehirn an die Wahrnehmung und an die Verarbeitung von Häufigkeiten adaptiert? *Beiträge zur 46 Tagung experimentell arbeitender Psychologen* Gießen, 2004, 134-138

Antell, S.E. and Keating, D.P. (1983). Perception of Numerical Invariance in Neonates. *Child Development, 54*, pp. 695-701.

Barsalou, L.(1999) Perceptual symbol systems. Behavioral and brain science. 22. 577-660

Church, R.M. and Meck, W.H. (1984). In H.L. Roitblat, T.G. Bever, and H.S. Terrace, Eds.: *Animal Cognition*. Erlbaum.

Corbetta, M, Miezin, F.M., Schulman, G.L., and Petersen, S.E. (1993). A PET Study of Visuospatial Attention. *Journal of Neuroscience, 13* (3), pp. 1202-1226.

Coren, S., and Porac, C. (1977). Fifty Centuries of Right-Handedness: The Historical Record. *Science, 198*, 631-632.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58,* 1-73.

DFG Projekt „Entscheidungsfindung unter Unsicherheit als fächerübergreifende Kompetenz − Alltagsorientierter Stochastikunterricht am Gymnasium (Ma

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, pp. 1-42.

Dehaene, S. (1997). *The Number Sense*. Oxford University Press.

Dehaene, S., Dehaene-Lambertz, and Cohen, L. (1998). Abstract Representations of Numbers in the Animal and Human Brain. *TINS, 21*, 8, pp. 355-361.

Fiddick, L. and Barrett, H.C. (2001). Evolution of Cognition: An Adaptationist Perspective. In: N.J. Smelser and P.B. Baltes (Eds.): *International encyclopedia of the social and behavioral sciences*. Vol. 7. Amsterdam: Elsevier, 2001, pp. 4996-5000.

Gallistel, C.R. and Gelman, R. (1992). Preverbal and Verbal Counting and Computation. *Cognition, 44*, pp. 43-74.

Gigerenzer, G. (2004). Calculated risks: What Numbers Really Tell Us About Our Lives. Simon and Schuster, New York.

Gigerenzer, P.M. Todd & the ABC Research Group (1999) Simple Heuristic that make us smart. New York: Oxford University Press.

Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684-704.

Goldstein, D., Hasher, L., and Stein, D.K. (1983). The Processing of Occurence Rate and Item Information by Children of Different Ages and Abilities. *American Journal of Psychology, 96*, pp. 229-241.

Hasher, L. and Chromiak, W. (1977). The Processing of Frequency Information: an Automatic Mechanism? *Journal of Verbal Learning and Verbal Behavior, 16*, pp. 173-184.

Hasher, L. and Zacks, R.T. (1979): Automatic and effortful processes in memory. *Journal of Experimental Psychologie: General, 108*, 356-388.

Hasher, L. and Zacks, R.T. (1984). Automatic Processing of Fundamental Information - The Case of Frequency of Occurrence. *American Psychologist, 39*, 12, pp. 1372-1388.

Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002). Representation facilitates reasoning: What natural frequencies are and what they are not. Cognition, 84, 343-352.

Kahneman, D., Slovic, P. & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological review, 80,* 237-251.

Kiefer, M. and Dehaene, S. (1997). The Time Course of Parietal Activation in Single-digit Multiplication: Evidence from Event-related Potentials. *Mathematical Cognition, 3*, 1, pp. 1-30.

Koerber, S. (2003) Visualisierung als Werkzeug im Mathematischen Unterricht: Der Einfluß externer Repräsentationen auf Proportionales Denken im Grundschulalter. Dr. Kovac Verlag.

Kurz-Milcke, E., & Martignon, L. (2005) Lebendige Urnen und Erlebnissreiche Bäume in der Grundschule. „Berichte des Arbeitskreises Stochastik in der Schule“ Nr. 3 Franzbecker: Hildesheim

Martignon, L. & Kurz-Milcke, E. (2005) Haufen bunter Steckwürfel für eine natürliche, gendersensitive Stochastik in der Grundschule Berichte des Arbeitskreises Stochastik in der Schule Nr. 3. Franzbecker: Hildesheim

Martignon, L.& Wassner, Ch. (2005) Schulung frühen stochastischen Denkens von Kindern Zeitschrift für Erziehungswissenschaft, 8. Jahrg., Heft 2, S.2002-222

Meck, W.H. and Church, R.M. (1983). A Mode Control Model of Counting and Timing Processes. *Journal of Experimental Psychology: Animal Behavior Processes, 9*, 3, pp. 320-334.

Osherson, D., Perani, D., Cappa, S., Schnur, T., Grassi, F., Fazio, F. (1999). Distinct Brain Loci in Deductive Versus Probabilistic Reasoning. *Neuropsychologica, 36*, 4, pp. 369-376.

Stern, E., Hardy, I. & Koerber, S. (2002) Die Nutzung graphisch-visueller Repräsentationsformen im Sachunterricht.In K. Spreckelsen, A. Hartinger, & K. Möller (Eds.) Ansätze und Methoden empririscher Forschung zum Sachunterricht (pp. 119-131). Bad Heilbrunn: Klinkhardt.

Wassner, C., Martignon, L., Biehler, R. (2004). Bayesianisches Denken in der Schule. In: Unterrichtswissenschaft. Zeitschrift für Lernforschung, 32. Jahrgang, Heft 1 (pp. 58-96). Juventa Verlag, Weinheim.

Zacks. R.T., Hasher, L., and Sanft, H. (1982). Automatic Encoding of Event Frequencies - Further Findings. *Journal of Experimental Psychology - Human Learning and Memory, 8* (2), pp. 106-116.

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, *98*, 287-308.

CONSTRUCTING STOCHASTIC SIMULATIONS WITH A COMPUTER TOOL - STUDENTS' COMPETENCIES AND DIFFICULTIES

Carmen Maxara and Rolf Biehler

University of Kassel, Germany

Constructing stochastic simulation with the computer tool Fathom™ has become an important part of an elementary stochastic course for student teachers at our university. We developed a three-step-design with a probabilistic part and activities with the software. Using a schema of the simulation process and different Fathom competencies we analyzed data from videotapes and semi-structured follow-up interviews. We sought to elaborate the problem solving process of students working on a simulation task focussing on how the students acquired Fathom and mathematical competencies. Some conclusions are that the competencies required by the students in different depths and the connection of Fathom objects to mathematical concepts have to be established in teaching.

INTRODUCTION

Simulation can serve two pedagogical purposes: Simulation can be used as a tool to solve problems and to make random situations more experiential (Biehler, 1991). Simulation is a mathematically more elementary method than calculating, so students can solve problems that are not otherwise solvable for them in a theoretical way. Also simulation can be used in combination or instead of analytical or combinatorial methods. We intend to use simulated models as experimental environments to support meaning construction in stochastic situations with various concepts like probability, event, random variable, or expected value. Another purpose for using simulation is to build up probabilistic intuitions.

At our university the tool-software Fathom is continuously used to support the learning and working processes in an elementary stochastic course for student teachers for grades 5 to 10 (pupils' ages are 11 to 16 years). The software is introduced and used for exploratory data analysis and descriptive statistics in the first part of the course. The students can use the learned software capabilities to analyze simulated data later on. The second part of the course concerns elementary probability. An important topic in this part is the simulation of random experiments, like the approach of Konold (1994) to estimate probabilities through simulation in introductory probability courses. The simulation in Fathom is introduced in parallel to the concepts of probability, random variables and events. Random situations are to be modelled mathematically and simulated; both results are to be compared if available.

The software is used as a student tool for actively analyzing data, simulating and building models as well as for exploring methods and concepts. During the whole term the students have to work with this software. The students are required to express the probability model in mathematical language, as well as using the language of events and random variables for explicitly expressing assumptions. The same requirements hold for student activities with Fathom. They have to construct a probability model in Fathom and they are to use the "Fathom-language" for defining events and random variables. Also, students have to use Fathom as a data analysis tool and to document their work and results.

THE STUDY

This study is part of a research program concerning teaching experiments with Fathom in elementary stochastic courses at our university as well as in "standard stochastic courses" at upper secondary levels at several schools (Biehler, 2003). Several studies aim to explore the relationship between students' stochastic thinking and knowledge and their simulation activities in Fathom in different ways (Biehler, 2006). Here we present some results of a case study with eight student teachers, who were asked to solve a simulation task with Fathom in pairs. We videotaped the working process and communication of the students and captured their computer activities with a screen capture program. Subsequently we watched the recorded material with every single student in individual sessions with a "method-mix" of stimulated recall and half structured interview.

Hypotheses of the study are based on three essential support features of the software Fathom for constructing simulations. One capability of Fathom allows students to construct and represent probabilistic models by various random machines. The random machines act for different kinds of random experiments. The second capability is the use of Fathom as a simulation tool itself. Theoretical probabilities can be estimated through relative frequencies and distributions of random variables through their empirical distributions. And third Fathom offers the possibility to experiment with the model.

To use Fathom for modeling, simulating and experimenting the students need certain Fathom competencies. Two questions of our research are therefore: Which Fathom competencies do students need? And to what extent did the students in our study acquire Fathom competencies to solve a typical task? A second field of research concerns the fact, that simulation with Fathom allows students to solve problems they would not be able to solve in a theoretical way, because in simulation mathematical competence is substituted by Fathom competence. Will thereby students more easily concentrate on probabilistic aspects during their problem solving process? Fathom reifies the concepts of event and random variable by means of the Fathom-object of a "measure": A random experiment is represented in Fathom as an attribute whose values are generated by a random machine (a random command). A “measure” is a function that can be defined on the results (the attribute values) of a random experiment. We also have the hypothesis that use of Fathom supports the use of these concepts as empirical concepts in the modelling context. But a transfer from empirical to theoretical terms probably works only with an additional theoretical treatment in the stochastic course (see the Summary and Consequences).

IDEAL WORKING PROCESS FOR SIMULATIONS WITH FATHOM - STAGES AND DECISIONS

In this section, probabilistic and simulation steps and decisions are described that take place in a normative task treatment. In a prior didactical analysis of the simulation capabilities of Fathom, we developed concepts and notions to use the software as a simulation tool. We illustrate the conceptualized activities first in an abstract way and second by a concrete task, and also show potential deviations from the ideal pattern and sources of problems. The following problem analysis is partly based on research results and on a theoretical analysis (Maxara & Biehler, 2006).

We envisage that students would work on simulation problems in three steps: setting up a stochastic model of the random situation by using probabilistic concepts, writing a plan of simulation and realizing the plan in Fathom. These three steps would be used as a modeling guideline for simulation.

In the first step the students are expected to model the random situation by building up a model of a real situation - a "model random experiment". They would describe it in a concrete model, for instance by an urn-model. They would construct the model, to identify the sample space, the probability distribution, as well as the events and random variables of interest. In the second step the students would transform the probabilistic concepts into a simulation plan. They could do this transformation step by step, perhaps as we exemplify it below.

For students' orientation we developed a four-step-design in two columns: probabilistic concepts and Fathom objects and operations, which correspond to each other in each step. We call the concrete description of the four stages in Fathom (right column) the "plan of simulation", built on ideas by Gnanadesikan, Scheaffer et al. (1987). The pedagogical intentions of the plan of simulation are that students structure their simulations, reflect about the simulations, and document their simulations.

Step	**Probabilistic concepts**	**Fathom objects & operations**
1	Construct the model, the random experiment	Choose type of simulation; define a (randomly generated) collection, simulate the random experiment
2	Identify events and random variables of interest (*Events and random variables as bridging concepts*)	Express events and random variables as “measures” of the collection
3	Repeat the model experiment and collect data on events and random variables	Collect measures and generate a new collection with values of the measures
4	Analyze data: relative frequency (events); empirical distribution (random variables)	Use Fathom as a data analysis software

Table 1: Four-step-design as a guideline for stochastic modeling

Our hypothesis is that the simulation plan is a helpful metacognitive tool for students and that Fathom is supportive of developing fluent simulation competence.

In the first step of the simulation plan one has to choose the type of simulation. We distinguish three different types of simulation in Fathom: the *simultaneous simulation*, in which the single experiment corresponds to different columns, the *sequential simulation*, in which the single experiment corresponds to different rows, and a *simulation as a sampling from an urn* (Maxara, 2006).

The four-step-design will be exemplified now on a typical task for the students. Thereby we demonstrate a perfect realization on the one hand and discuss possible problems on the other hand.

The problem: Mister Becker has to wear a black suit during his working hours, but he can choose the tie himself. 7 different ties hang in his wardrobe. Every morning he randomly takes one tie out of the wardrobe and puts it back in the evening.

1. What is the probability that Mister Becker will wear 5 different ties in his five-day working week?
2. What is the probability that Mister Becker will wear at least two identical ties in his five-day working week?
3. How many different ties does Mister Becker wear on average in his five-day working week?

Tasks:

a) Formulate the situation as a compound random experiment and specify the sample space W and the probabilistic assumptions.
b) Provide a plan of simulation and estimate the unknown probabilities of 1., 2., and 3 with a Fathom-simulation.
c) Document your work in a Fathom file.

Stage 1: Defining the random experiment

Perfect realization: On the probabilistic side the students have to make the following assumptions: each tie has the same probability to be drawn and the samples are stochastically independent. The model could be a sampling from an urn that contains seven different balls, and five balls are successively sampled with replacement. In Fathom they have to choose an appropriate type of simulation. In this case this could be the sequential simulation, or sampling from an urn. We choose the sequential simulation. There we have to define a collection (e.g. named "week") and an attribute "tie" with the random machine "randomInteger(1;7)" or "randomPick(1;2;3;4;5;6;7)". Finally, we have to add five cases to the collection. Fathom is a supportive tool because it offers different random machines that resemble to concrete models.

Possible problems: During the construction of a simulation several decisions have to be taken. Each decision one has to take is a potential source of error with for a larger or smaller impact on the simulation and the interpretation of the simulation. In our study we could identify the following two problems: The first problem area is the transformation of the random experiment into a correct simulation in Fathom. This problem depends on decisions one, four and five. The second problem is the naming of the objects (decisions two and three). The decisions of naming are important as the names of objects are relevant for the interpretation of results at a later stage.

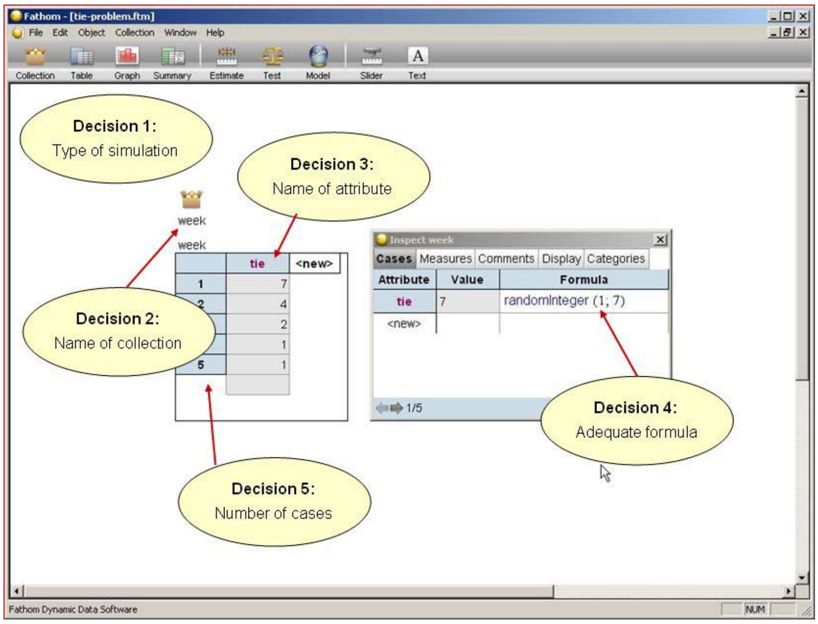


Figure 1: Sequential simulation of the tie-problem with 5 decision points

Stage 2: Events and random variables - defining measures

Perfect realization: In the second step, events and random variables should be identified and defined. On the probabilistic side they should be verbalized. For this example, we can define the two events by E1: "Mr. Becker wears five different ties in a week", E2: "Mr. Becker wears at least two equal ties in a week", and for the expected value we have to define the random variable X: "Number of different ties in a week". In Fathom, events and random variables correspond to *measures* that refer to the collection as a whole. In the collection, we have to designate and define three measures with adequate formulas: For E1: "five_different_ties": *uniqueValues(tie)=5*, for E2: "least_2_equal_ties": *uniqueValues(tie)* $\leq$ *4*, and for X: "number_different_ties": *uniqueValues(tie)*. Fathom is supportive in two respects: the concept of "measure" is a natural representation of an event or random variable, and second, commands such as "uniqueValues" makes this typical type of analysis direct and easy.

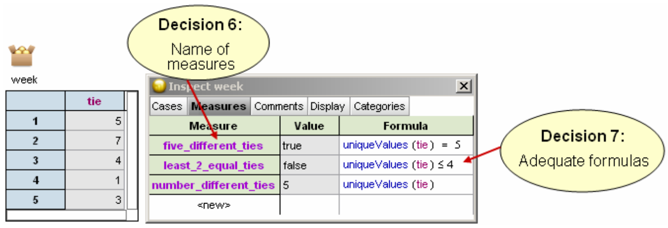


Figure 2: Defining measures

Possible problems: The problems in this stage are the naming of the measures (decision six) and the knowledge and implementation of correct formulas (decision seven). Another identified problem is that students omit the description and distinction of events and random variables before defining measures, and they try to transform (only) their colloquial verbalization into measures.

Stage 3: Repeating the (compound) experiment - collecting measures

Perfect realization: The model experiment has to be repeated. Thereby we have to collect data about events and random variables and to decide on the number of repetitions. In Fathom, this corresponds to collecting *measures*. If you collect *measures* in Fathom a new collection of measures is provided with five automatic repetitions. Now, you have to collect as many measures as you think are adequate for approximately estimating probabilities or distributions. In this example, the simulation was repeated 1000 times.

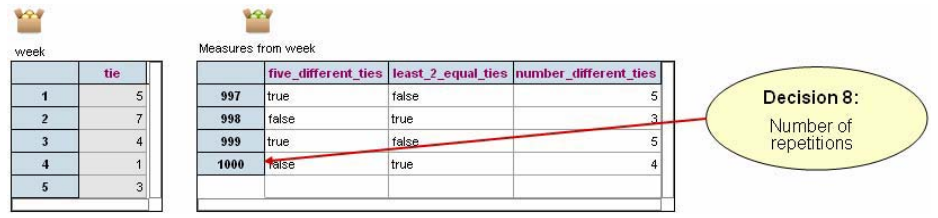


Figure 3: Collecting measures - repeating the simulation

Possible Problems: One problem field is the decision of the number *n* of repetitions. Are students reflecting about *n* or do they always take 1000 repetitions by standard practice?

Stage 4: Data analysis

Perfect realization: The fourth step consists of data analysis. The probabilities that were asked for in questions one and two can be estimated by the relative frequencies of the respective event. The expected value can be estimated by the empirical mean of the random variable. In Fathom, we can use summary tables to calculate the empirical values by the formulas "columnproportion" for the relative frequencies of the events and "mean( )" for the mean of the random variable. The graphics (bar chart and histogram) can be used to visualize the data, but also to read off the values by dragging the cursor on the bars.

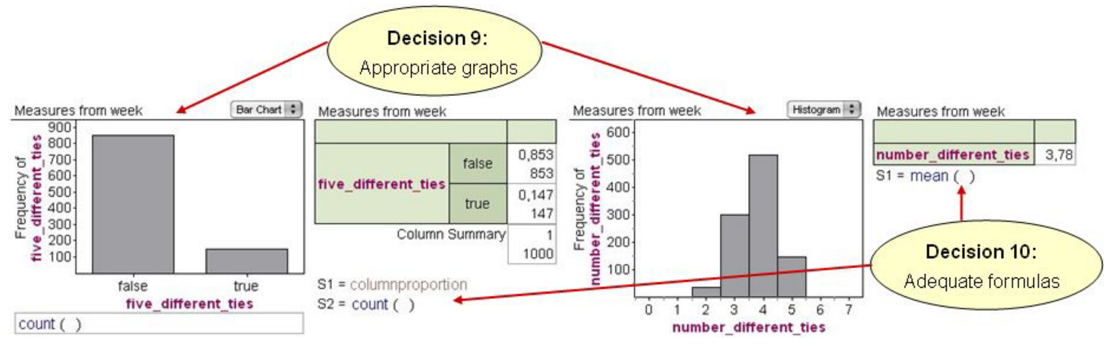


Figure 4: Data analysis in Fathom

Possible Problems: One source of problem is how to technically find the distribution and relative frequencies with Fathom. Another type of problem is related to interpreting the computed values. For instance, a mean of 3.78 could be rounded to 4 (ties per week), showing by this a misunderstanding of means and expected values.

STUDENTS' COMPETENCIES AND DIFFICULTIES DURING THEIR SIMULATION ACTIVITIES

The students have to have different stochastic and Fathom competencies to cope with the simulation in Fathom and the possible problems they might encounter. Our working group distinguished four Fathom competencies: *general Fathom competence*, *formula competence*, *simulation competence* and *strategic and generalizing competencies* (see Keitzer (2006) for a first application of these concepts). The general Fathom competence contains the knowledge of tools and their basic functionality in Fathom as well as the handling of the objects on the screen (clear design). The handling of the formula editor and the knowledge of formulas in different contexts were integrated into formula competencies. The simulation competence describes the competence to transform the random experiment into a Fathom simulation and to give meaningful naming of objects. The strategic and generalizing competencies catch the handling and avoidance of errors.

We have used these Fathom competencies and the introduced schema of the simulation procedure to analyze the problem solving activities of the eight students. The possible problems mentioned in each stage of simulation needed different Fathom competencies to be handled: Problems of transformation require simulation and formula competence, problems of naming are part of simulation competencies, and knowledge about formulas is integrated into the formula competence. But not all problems are only connected with Fathom competencies, they also involve probabilistic competencies, reflection and the ability to connect the two sides. Those problems are for instance, the definition and distinction of events and random variables and the interpretation of the computed results. Below, we illustrate some exemplary competencies and difficulties of the students.

Exemplary competencies

All pairs were able to solve the task through a simulation in Fathom and obtained correct results. Students' competencies were found in all four distinguished Fathom competencies and also on the probabilistic side.

Simulation competence: To illustrate the simulation competence we look at the following part of transcript.

1 S 1: I shall now, well, we have chosen the sampling simulation.
2 S 2: Mmh.
3 S 1: We have to put seven ties into the collection.
4 S 2: Mmh.
5 S 1: And then I would take a sample …
6 S 2: Mmh.

7 S 1: … of size five.
8 S 2: Mmh.
9 S 1: Wouldn't I? Because it refers always to one week.

Student 1 explains to the other student what to do next and what the Fathom steps mean in terms of the random experiment. She has a reflective view on their working and simulation process and makes up a plan of the following simulation steps.

Formula competence: The following example will show a good formula competence. Right at the beginning to simulate the random experiment the students have created a collection with an attribute.

10 S 3: And these are the ties. (*Names the attribute "ties"*)
11 S 4: Mmh.
12 S 3: And these are -, how many are in there?
13 S 4: Of what? We have seven ties, haven't we? (*Opens the formula editor*) RandomInteger. (*Type the formula*)
14 S 3: One comma seven.

The students do not talk about which formula they have to use, one of them stating the correct formula and the other completing it with the required values. It seems that they have adapted the random machines as representations of models for random experiments and do not have to think much about an adequate formula.

Stochastic competence: The following example shows students' understanding of the concept of the law of large numbers. A pair talked about the approximation of the relative frequencies of the measures to the theoretical probabilities and wrote into their document: "One can regard the relative frequencies of the measures as an approximation for the probabilities. The probability that Mister Becker wears five different ties is approximately 16.6%." By this sentence the students show an understanding of the estimation aspect, the fact that the simulated result is an approximation of the theoretical probability and not the probability itself.

Exemplary difficulties

We identified several domains of difficulties during the four stages: difficulties in transforming the random experiment into a simulation, lacks in formula competence and difficulties pertaining to the probabilistic part and the connection to the Fathom simulation. Here are only some examples for the observed difficulties.

Problem: Omitting probabilistic steps After defining the random experiment as a simulation, students omit the description and distinction of events and random variables and transform their colloquial verbalization or the task itself into a measure. In this case the students have created the simulation of the random experiment without discussion about what events or random variables to define as measures.

32 S 3: And now we should define measures. How do we call the first measures? (*She opens the inspector of the collection*). Anyway we have more then one. (*Naming the first measure "E1"*)

33 S 4: Mmh. (*S3 types the formula uniqueValues(tie)=5*)
34 S 4: Yes?
35 S 3: Yes.

The naming of the measure indicates that the students do not think about the mathematical type of object. The reification of event and random variable in Fathom is not strengthened as intended in the theoretical part of the course. These two concepts are rather blurred in students' minds, probably because both of them were defined as measures. The students use the Fathom-concept of measures as a replacement for the stochastic concepts of event and random variable in an efficient way. This shows a need for additional theoretical concept building.

Difficulty: Interpreting the results Difficulties in interpreting simulated results are related to a still unsatisfactory comprehension of the difference between relative frequency and probability. The simulated relative frequency of 0,171 is interpreted (at least in the text) as the (definite) probability to use five different ties per week. Here the students do not clearly distinguish consciously between those two concepts.

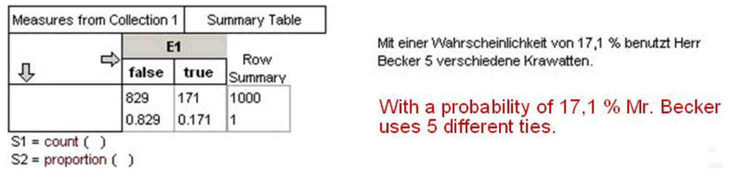


Figure 5: Exemplary interpretation of a simulated result

SUMMARY AND CONSEQUENCES

Despite some difficulties, the students of our study were able to simulate these kinds of stochastic situations in Fathom. They acquired the intended software competence and accepted the plan of simulation as guidance. We have distinguished four Fathom competencies that the students should have to simulate such kinds of random experiments. The general Fathom competence was acquired by all students. The other three competencies were acquired in different depths like the examples illustrate.

The substitution of mathematical competence by simulation competence is not difficult for the students. But most students do not use this freedom to concentrate on the probabilistic concepts behind the simulation. The reification of event and random variable as different kinds of measures did not work as expected. These students do not distinguish explicitly between events and random variables during their simulation process, but they have a working knowledge about them. The concepts tend to remain separated in the two worlds, the "World of probability" and the "World of Fathom". Thus – as an instructional implication of the study – the informal use of these two concepts should be more deeply and explicitly related to each other in the course.

As a consequence it looks expedient to put more emphasis on the explanatory aspect of the simulation plan, so as not to only provide technical help, but also to foster students' language, knowledge and reflection to link simulation aspects with probabilistic concepts. Students should reflect on their simulation activities, about the basic probabilistic concepts and their relation to the simulation. Another aspect is to support a closer relation between the "World of probability" and the "World of Fathom" through more comparisons of simulations and theoretical mathematical solutions.

The conclusions of this paper are also relevant to other countries than Germany in different aspects, because both the understanding of probability concepts and the simulation of random experiments are essential in teaching statistics. The use of technology is an important aspect in teaching and learning statistics. We think that our modelling guideline for simulation – setting up a stochastic model, writing a plan of simulation and realizing the plan in Fathom – is adaptable (perhaps with some modifications in relation to the range of random experiments) for other simulation-software. The general categories of students' competencies and difficulties are also applicable to other software, thus they could be of general interest to statistics instructors and researchers.

REFERENCES

Biehler, R. (1991). Computers in probability education. (In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: Probability in education* (pp. 169-212). Dordrecht.)

Biehler, R. (2003). *Interrelated learning and working environments for supporting the use of computer tools in introductory classes.* (Paper presented at the IASE satellite conference on Statistics Education, Berlin. http://www.stat.auckland.ac.nz/~iase/publications/6/Biehler.pdf)

Biehler, R. (2006). *Working Styles and Obstacles: Computer-Supported Collaborative Learning in Statistics.* (Paper presented at the ICOTS 7, Salvador da Bahia, Salvador. http://www.stat.auckland.ac.nz/~iase/publications/17/2D2_BIEH.pdf)

Gnanadesikan, M., Schaefffer, R. L. & Swift, J. (1987). *The Art and Techniques of Simulation*. (Palo Alto, CA: Dale Seymore.)

Keitzer, C. (2006). Selbständig-kooperative Bearbeitung von stochastischen Simulationsaufgaben am Computer - Qualitative Analysen zu Schülerkompetenzen und Arbeitsweisen. Diplomarbeit, Universität Kassel

Konold, C. (1994). Teaching Probability through Modeling Real Problems. *The Mathematics Teacher, 87*(4), 232-235.

Maxara, C. (2006). *Einführung in die Stochastische Simulation mit Fathom*. Kassel: Universität Kassel. http://nbn-resolving.org/urn:nbn:de:hebis:34-2006082514477

Maxara, C. & Biehler, R. (2006). *Students' Probabilistic Simulation and Modeling Competence after a Computer-Intensive Elementary Course in Statistics and Probability.* (Paper presented at the ICoTS 7, Salvador da Bahia, Brazil. http://www.stat.auckland.ac.nz/~iase/publications/17/7C1_MAXA.pdf)

Fathom Dynamic Statistics^{TM} *Software*. Emeryville, CA: Key Curriculum Press.

A CROSS-NATIONAL COMPARISON OF INTRODUCTORY STATISTICS STUDENTS' PRIOR KNOWLEDGE OF GRAPHS

Maria Meletiou-Mavrotheris* and Carl Lee**

*Cyprus College, Cyprus, **Central Michigan University, USA

This study investigated the prior knowledge about graphing that groups of undergraduate Cypriot and U.S. students brought into the introductory statistics classroom. A total of 159 students completed a questionnaire designed to assess three aspects of graph comprehension: graph reading and interpretation, graph construction and graph evaluation. The study findings confirm our initial conjecture that U.S. students would exhibit better graphing skills due to the higher emphasis on statistics in U.S. school mathematics curricula. U.S. students outperformed their Cypriot counterparts in all tasks. The biggest differences, however, were observed in simple reading and interpretation tasks. Both Cypriot and U.S. students had difficulties in tackling more demanding tasks involving group comparison, graph construction, and critical evaluation of information presented graphically.

INTRODUCTION

Introductory statistics courses have been using a multitude of graphical representations both as an essential tool in statistical investigation, and as a means to communicate statistical ideas. Good graphing skills are essential to conducting meaningful data analysis. Hence, it is important to know, at the outset of instruction, what types of data representations students are familiar with, and what difficulties they might encounter in the construction, interpretation, and evaluation of graphs.

This article presents findings from a study that compared the background knowledge about graphs of undergraduate Cypriot and U.S. students upon entering an introductory statistics course. We chose to compare US and Cypriot students because they come from two educational systems that put different emphasis on statistics at the school level. In the U.S., like in many other countries, statistics has been established as a vital part of school mathematics at all grade levels. In Cyprus, while statistics spans the elementary school mathematics curriculum, there is almost complete absence of statistical concepts from the secondary school curriculum. Given the higher exposure of US students to statistics while at school, we anticipated that they would likely exhibit better graphing skills than their Cypriot counterparts. On the other hand, since charts, graphs, and plots are used broadly in the media to present, disseminate and "explain" information (Shaughnessy, Garfield, & Greer, 1996), we conjectured that college-level Cypriot students would still be familiar with the main types of graphs despite less extensive formal study at school.

METHODOLOGY

Context and Participants: The sites for the study were five introductory statistics courses across three campuses – two four-year colleges in Cyprus, and a Midwestern university in the United States. A total of 159 students (92 Cypriot students, 67 U.S. students) participated in the study. Most students were sophomores or juniors and majored in Business. Few had taken mathematics courses at the precalculus level or higher. The average age for both US and Cypriot students was around 20. Only a very small percentage of the students were adult learners.

Instruments, Data Collection and Analysis Procedures: At the beginning of the semester, students were administered a questionnaire. An identical questionnaire was administered to both US and Cypriot students, as the language of instruction in the Cypriot institutions participating in the study is English. The questionnaire consisted of ten tasks, covering four basic types of statistical graphs: bar graph, pie chart, line plot, and time plot. It was designed to assess three aspects of graphical knowledge: graph reading and interpretation, graph construction, and graph evaluation. Each of the tasks was selected from previous studies in statistics education to provide a point of reference for our findings. The tasks were open-ended, requiring students to justify their responses. Four of the tasks are shown in Figure 1.

Student responses to the questionnaire sub-tasks were first grouped into three categories depending on the aspect of graphical knowledge that was being assessed, and were then analyzed using the constant comparison analysis method. Constant comparison analysis, which involves unitizing, categorizing, chunking, and coding by choosing words, phrases, or sentences that specifically address the research questions, assisted us in the search for patterns and themes that were used to develop the study's interpretation. We reached closure only after many sweeps through the data. The main findings of this empirical analysis are outlined in the next section.

RESULTS

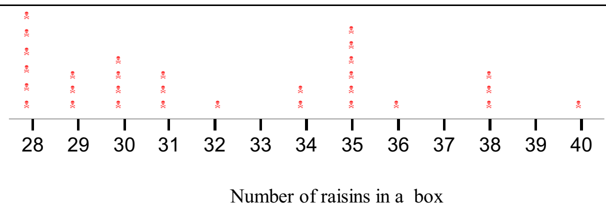
A number of learning patterns and trends were discerned by examining U.S. and Cypriot students' responses to the questionnaire. These patterns are described and interpreted for each of the three components of graph knowledge investigated in this study: graph reading and interpretation, graph construction, and graph evaluation.

Reading and Interpretation

Graph interpretation involves forming opinions from one or more graphs. It includes making comparisons within or between data sets displayed graphically, identifying patterns and trends, and making inferences from graphs. Graph interpretation cannot be effective if the reader does not possess basic graph reading skills. Thus, readers' interpretations of a graph provide, at the same time, evidence of their knowledge of the graph's structure (Friel, Curcio, and Bright, 2001).

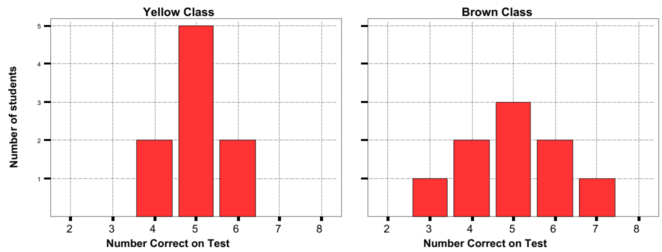
Most of the Cypriot participants did very poorly on the tasks which explored their graph reading and interpretation skills. Their performance was particularly low in tasks involving reading and interpretation of bar graphs; they seemed unsure as to what the

Task 1: "Raisins in a Box" (Friel & Bright, 1996)



Are there the same number of raisins in each box? How can you tell?

Task 2: "Students' Scores" (Watson & Moritz, 1999)



Which class did better?

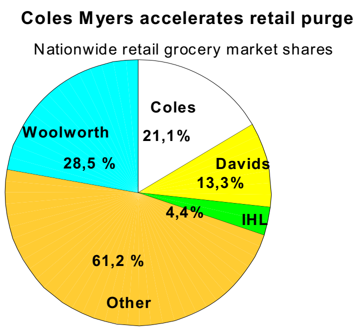
Task 3: "Top Actors' and Actresses' Salaries" (Jones et. al, 2002)

Salaries of 15 Top Actors and Actresses (in millions of dollars)	
Actors	*Actresses*
$17.5	$12.5
15.0	9.0
20.0	11.0
20.0	9.5
20.0	2.5
19.0	12.0
20.0	3.0
18.0	4.0
5.5	4.0
6.0	2.5
10.0	6.0
16.5	8.5
12.5	4.5
10.0	3.0
7.0	10.0

c) Construct a graph that will allow you to compare the salaries of actors and actresses.

d) How do the actors' salaries compare to the actresses' salaries?

Task 4: "Grocery Market Shares" (Watson, 1997)



a) Explain the meaning of this pie chart, which appeared on Australian Financial Review (1993).

b) Is there anything unusual about it?

Figure 1: Sample of Questionnaire Tasks

Type of response	Cypriot Students		U.S. Students	
	n	*%*	*n*	*%*
1. Properties of the graph – considering both the range of data and frequency "No, there aren't the same number of raisins in each box you can tell because there are a different number of students with different numbers of raisins in a box. The boxes have 28 to 40 raisins inside, so they aren't the same."	6	7	16	24
2. Literally "reading" the data from the graph "No. In 6 boxes there are 28 and in 6 other 35. In 3 boxes there are 29 and in other 3 31 and 38. In 4 there are 30, in 2 there are 34 and in 1 there are 32, 36, and 40."	5	5	4	6
3. Range of the data – considering only range and does not include frequency "No, the number of raisins in each box varies from 28-40 per box."	9	10	40	60
4. Frequency of occurrence/height of bar "No, there aren't. In some boxes there are many raisins and in some others no raisins at all. In few boxes you can see that is the same number of raisins."	19	21		0
5. Properties related to the context or to the data "No, because they found different numbers because weight is not exact as number"			1	1
6. Other - incomplete, unclear, or not statistically reasoned responses "No, there is not the same number of raisins in each box."	23	25	4	6
7. No response/I don't know	30	32	2	3
	92	*100*	*67*	*100*

Table 1: Patterns of Cypriot and U.S. students' responses to Task 1

axes of bar graphs represent and had difficulties in distinguishing between data values and frequencies. They also had difficulties in reading simple plots of raw data. The only graphs they were somewhat familiar with were the pie chart and the time plot. U.S. students, on the other hand, gave responses indicating familiarity with all four types of graphs included in the questionnaire (bar graph, pie chart, line plot, and time plot), and possession of basic graph reading and interpretation skills.

Cypriot students' responses to Task 1 (see Figure 1) are indicative of their lack of familiarity with even very simple graphs such as the time plot. In this task, students were given a line plot depicting the quantity of raisins in half-ounce boxes, and were asked to determine whether all boxes had the same number of raisins. In analyzing student responses, we used the same coding scheme as that used by Friel and Bright (1996), who included the task in a study researching middle school students' level of graph comprehension. The results of the analysis are displayed in Table 1, which shows the frequency and proportion of each pattern of responses among U.S. and Cypriot students, as well as examples of typical responses within each pattern.

Data in line plots are ungrouped, making them easier to interpret than bar graphs or histograms. However, despite the seemingly easy nature of Task 1, Cypriot students did extremely poorly. More than half (57%) of them gave either no response or some incomplete or vague response, while only about a fifth (22%) gave a reasonable response (Response types 1-3 in Table 1). By contrast, almost all U.S. students (90%) gave reasonable responses. And while a sizeable proportion of Cypriot students (21%) confused the role of data values and frequencies and focused on the frequency

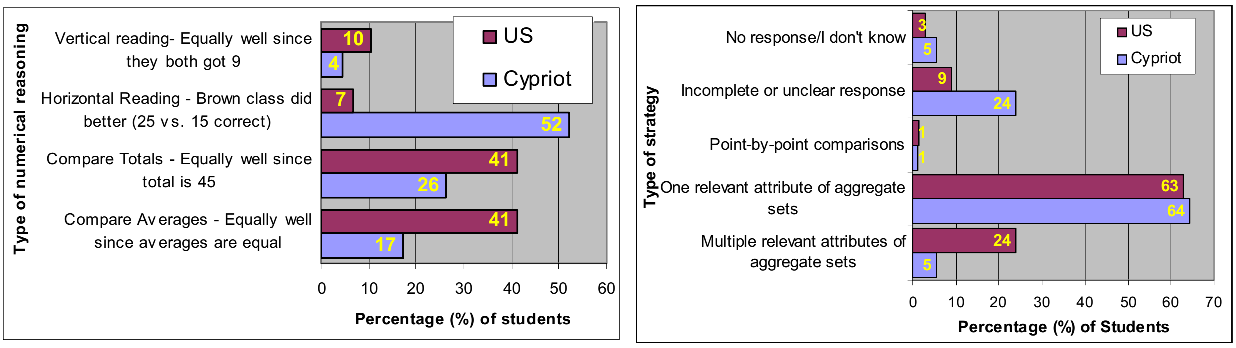


Figure 2: Patterns of Cypriot and U.S. students' responses to Task 2

or number of X's as the data values themselves (Response Type 4 in Table 1), no U.S. student did so. Although definitely exhibiting better graph interpretation skills than Cypriot students, students from the U.S. gave responses which indicated that they focused on only one feature of the distribution displayed by the line plot. The most common response among U.S. students was Type 3 response, i.e. considering only the range of values included in the line plot. Only a fourth of the U.S. students (24%) considered both the range of data and the frequency of occurrence.

The comparison of univariate datasets displayed graphically is a graph interpretation activity that has been investigated in several studies of secondary school students. Task 2 was included in the questionnaire to assess this important data analysis skill that provides the foundation to more formal group comparisons using inferential statistics (Watson and Moritz, 1999). Students had to compare the scores on a test of two classes of equal size based on a graphical representation of the scores. We coded student responses using the following pattern descriptors: (1) Comparing datasets by using multiple relevant attributes of the aggregate sets; (2) comparing datasets by using one relevant attribute of the aggregate sets (e.g. shape, center, or spread); (3) comparing datasets by using a point-by-point strategy (Groth, 2003).

For this task also, there was a noticeable difference between the responses given by Cypriot and U.S. participants, with U.S. students again outperforming Cypriots. However, only a small proportion of U.S. students showed evidence of a powerful sense of distributional reasoning in group comparison problems. As seen in Figure 2-left, only one-fourth of U.S. students (24%), and an even smaller percent of Cypriot students (5%), exhibited the first, most sophisticated pattern of responses in Task 2. These students viewed the two sets of scores displayed in the graphs as aggregates and used two or more relevant characteristics (e.g. center and spread) to make their comparisons: *"Brown b/c more kids got higher than six. In Yellow Class, they were all in the middle"; "They did about the same on average, but Yellow Class was more consistent"; "Yellow Class because even though none of them got 7 right, they all were average, none got 3 right like Brown Class".*

Two-thirds of both Cypriot and U.S. students exhibited the second pattern of responses. They compared the two datasets based on a single feature of the aggregate sets. The feature used most often as the sole criterion of comparison, was the center of the two distributions: *"Classes scored the same. If you take the class average, you'd get the same for both Yellow Class and Brown Class"*. It should be pointed out here that the nature of the question posed to students in this task, might have contributed to this observed tendency to focus only on the center of the data distributions. Since students were only asked to decide which of the two classes did better, they might have concluded that comparing centers would be adequate, and that making reference to spread would not further enhance their answer.

The least sophisticated pattern of responses involved a point-by-point comparison of the two sets. **"***Yellow. Although Brown had a 7 on a test, they also had a 3, while Yellow had no 3s but 2 more with 5s, eve though no 7s"; "They were equal because Brown had a student that got more correct. Yellow had more students get more correct on a certain number"*. Only two students exhibited this type of reasoning.

Finally, almost a third of the Cypriot students (29%), in contrast to only twelve-percent of U.S. students, either gave no response, or some incomplete or unclear response that made it impossible to classify their reasoning strategy: *"It is the same because they both have correct answers on the test"*; *"Yellow Class did better"*.

In analyzing several tasks, including Task 2, we investigated whether students approached the tasks visually, numerically, or using a combination of visual and numerical techniques (Watson and Moritz, 1999). The most popular approach was the numerical one – perceiving graphs as a way to obtain the actual scores in order to calculate a number that would summarize the data. In Task 2, more than half of both the U.S. and Cypriot students used only numerical techniques. We identified four types of numerical strategies in tackling Task 2 (see Figure 2-right). The big majority of U.S. students (82%) used either the "compare averages", or the "compare totals" strategy. Both of these numerical strategies are valid in the context of this particular task, although using totals as a basis of comparison would not have been correct had the classes not been of equal size. The most popular numerical strategy for Cypriot students, on the other hand, was the "horizontal" reading strategy (focusing on the horizontal axis and comparing only values). Both the "horizontal" reading strategy, and the "vertical" reading strategy (comparing only heights of bars), suggest limited understanding of the function of the axes in a bar graph (Ben-Zvi, 2004).

Graph Construction

Graph construction is the process of displaying one or more datasets by using graphs. In order for students to be able to construct effective graphical representations of data, they need to know how to organize data, understand graph conventions like scaling and labeling axes, but, more importantly, also know which graph is the optimal choice for a given situation (Friel, Curcio, and Bright, 2001).

The graph construction tasks included in the assessment required students to draw their own graphs of either provided or projected data. We did not specify the kind of graph students should construct because we were interested in seeing whether they would produce a graphical representation that conveyed the different characteristics of a data distribution (center, spread, shape). Thus, in examining student constructions, our main focus was not on the technical details of their graphs, but on the characteristics of the distribution shown in the graphs, and particularly on whether the plot revealed or masked data variation (Canada, 2004). Of course, the characteristics of the underlying distribution can only be assessed when the technical details of the produced graph are plausible or at least understandable (Canada, 2004). Also, the type of graph that students use depends on their repertoire of different graph types, and on their skills in drawing graphs. Thus, while our focus was on assessing students' distributional reasoning as judged by the graphs they constructed, we also paid some attention to the technical details of their graphs.

The questionnaire tasks assessing students' ability to construct graphs proved quite challenging for most students. Although students from Cyprus had a notably lower performance than U.S. students in graph construction tasks also, U.S. students did not perform well either. In Task 3, for example, where students had to construct a graph that would allow them to compare the salaries of fifteen top actors and actresses shown in a table, more than half of the Cypriot students (54%), and one-third of the U.S. students (33%), either gave no response or drew an empty plot, i.e. a plot with no data values in it (see Figure 3-left). Only thirty percent of U.S. students and fifteen percent of Cypriot students managed to construct a valid display that revealed the underlying distributions of salaries and allowed comparisons (e.g. integrated data into a single display that used ranks). Another twenty-two percent of U.S. students and eighteen percent of Cypriot students attempted to construct a graphical form that would provide information about both the center and spread of the two distributions of salaries, the displays they produced, however, were not valid and did not allow for proper comparisons between the two sets of salaries. Also, some students (15% of U.S. students, 13% of Cypriot students) produced a graph that displayed a single aspect of the data (e.g. a value bar with only two values – the first from each category, or a bar graph of mean salaries for actors and actresses).

Graph Evaluation

Graph evaluation is the ability to look behind the data rather than simply accepting the initial impression given by a graph (Monteiro & Ainley, 2003). It involves evaluating a graph on its correctness or effectiveness. As Watson (1997) points out, statistics instruction in the high school years should aim at gradually building students' ability to question unrealistic claims made by the media or others without proper statistical foundation. Task 4 (see Figure 1) assessed not only the ability to read and interpret a graphical form (the pie chart), but also the ability to critically evaluate and question information presented graphically. Watson (1997) included this task in a large survey administered to middle school students in Australia to assess their level of statistical thinking based on authentic extracts from the media.

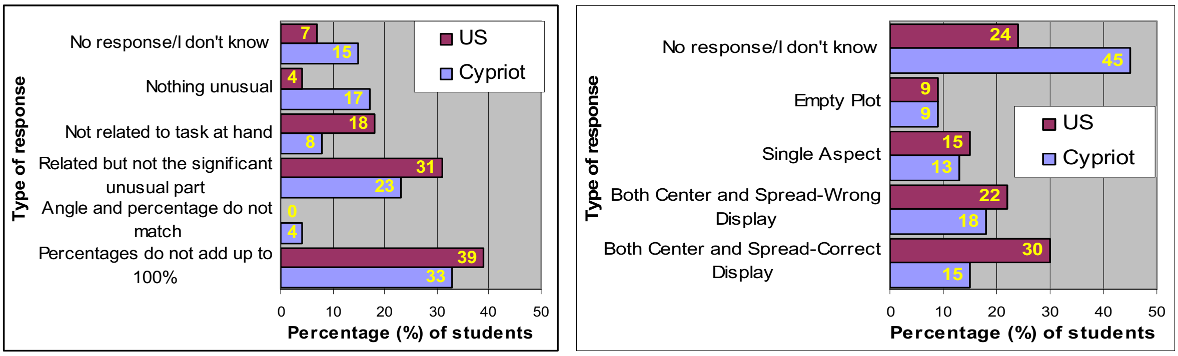


Figure 3: Patterns of Cypriot and U.S. students' responses to Task 3(Figure 3-left), and Task 4 (Figure 3-right)

In the first part, where students had to explain the meaning of the pie chart (i.e. a reading and interpretation task), the majority of both Cypriot and U.S. students gave acceptable responses indicating basic understanding of the pie chart. In the second part, however, where they had to recognize the error in the pie chart, results were discouraging for both groups. Only thirty-nine percent of U.S. students and thirty-seven percent of Cypriot students recognized that the percentage figures given in the pie chart are incorrect (Figure 3-right). These students either pointed out that *"the percentages do not add up to 100%"*, or noted that *"angle and percentage do not match"*. One third of Cypriot students (32%) gave no response, or concluded that there is nothing unusual with the pie chart. A higher proportion of U.S. students (49% of U.S. students vs. 31% of Cypriot students) realized that there was something wrong with the pie chart, but gave either an explanation not being related to the task at hand, or an explanation being related but not being the significant unusual part to one who fully understands the structure of a pie chart (Watson, 1997). Explanations not related to the task at hand focused on the appearance of the graph, rather than on the message the graph meant to convey: *"There is no key with the different colors and explaining what things mean. There should be decimal points"*.

Among responses related to the task but not being the significant unusual part, the most common one was in the spirit of the following: *"The other section (people that produce too small an amount to be named individually) fills up more than half of the market shares"*. This peculiar feature of the pie-chart in the task attracted the attention of many students. The proportion of both US and Cypriot students displaying this type of reasoning was very high. This might be indicative of a limitation of the specific item, since indeed "*Other being bigger than the rest*" is something that one would never encounter in a valid pie-chart displaying real data.

DISCUSSION

In this study, we compared the background knowledge about graphs of groups of introductory statistics students in Cyprus and the U.S. The study findings confirm our initial projection that U.S. students would exhibit better graphing skills. U.S. outperformed their Cypriot counterparts in all tasks. The biggest differences were observed in graph reading and interpretation tasks. U.S. students gave responses that indicated they had developed the ability to read and interpret basic plot types, in contrast to Cypriot students, the majority of whom exhibited lack of basic graph reading skills and serious misconceptions about graphs. However, both student groups exhibited limited distributional reasoning. They tended to focus on one feature of data distributions – usually the center – rather than considering multiple relevant attributes (shape, spread, skewness, etc.). Also, both student groups did poorly on tasks requiring them to critically evaluate information presented graphically.

This study was exploratory in nature. There are several methodological weaknesses that might limit the validity and generalizability of the findings. The small scale and limited geographical nature means that generalizations should be done cautiously as the specific classrooms investigated might not be representative of all introductory statistics classrooms in the US and Cyprus. Moreover, the fact that graphical knowledge was solely assessed using a questionnaire, means results are limited by the appropriateness of the test items included in the questionnaire. One should also recognize the limitations imposed by language. Naturally, Cypriot students' ability to express themselves in writing was typically lower than that of native speakers, and this might have somewhat contributed to their lower performance.

Clearly, the results presented here are only suggestive and warrant more rigorous study. To better understand cross-national differences in students' graphical knowledge, there ought to be triangulation of data, using a combination of qualitative and quantitative means of data collection. Also, one ought to draw attention to the underlying cultural factors affecting statistical achievement. She/he has to consider the unique culture and characteristics of each country rather than assuming that instructional variables operate in the same way across all countries. She/he should also relate differences in student achievement to the broader educational system characterizing the two nations: coverage of topics related to statistics in the intended and implemented curriculum, academic training of teachers in statistical content and pedagogy, teachers' beliefs about the nature of statistics, technology availability and use in the classroom, assessment practices, trends in education reform, etc.

Despite its weaknesses, the study does provide useful information regarding college-level students' background knowledge of graphs. Findings corroborate with the research literature which indicates that most college-level students across nations lack a global perception of the features of a data distribution displayed graphically (e.g. Ben-Zvi, 2004), and have little understanding of data beyond simple bar-charts and pie-charts (e.g. Rubin, 2002). Moreover, findings suggest marked differences in the level of background graphical knowledge among college-level students from different countries that might be related to variations in the school mathematics curricula. Cypriot students' poorer knowledge of graphs compared to U.S. students can partially be attributed to their more limited exposure to statistics while at school.

The expanding use of data for prediction and decision-making in almost all domains of life makes it a priority for educational systems worldwide to help all students develop understanding of key statistical ideas prior to entering college. Although we by no means suggest imposing an international statistics curriculum, we do believe that statistics should be established as a vital part of school mathematics in every country. This is necessary in modern, knowledge-based society, where the ability to analyze, interpret and communicate information from data are skills needed for daily life and effective citizenship.

REFERENCES

Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal, 3*(2), 42-63.

Canada, D. (2004). *Preservice teachers' understanding of variation.* Unpublished doctoral dissertation, Portland State University, Portland, Oregon (USA).

Friel, S. N., and Bright, G. W. (1996, April). *Building a theory of graphicacy: How do students read graphs?* Paper presented at the annual meeting of the American Educational Research Association, New York, NY

Friel, S. N., Curcio, F. R., Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, *32*, 124-158.

Groth, R. E. (2003). *Development of a high school statistical thinking framework.* Unpublished doctoral dissertation, Illinois State University, Illinois (USA).

Jones, G., Mooney, E., Langral C., and Thorton, C. (2002). Student's individual and collective statistical thinking. *Proceeding of the Sixth International Conference on teaching Statistics* [CD ROM]. Durban, South Africa.

Monteiro, C., and Ainley, J. (2003). Interpretation of graphs: Reading through the data. In J. Williams (Ed), *Proceedings of the British Society for Research into Learning Mathematics, 23* (3), 61-66.

Rubin, A. (2002). Interactive Visualizations of Statistical Relationships: What Do We Gain? *Proceedings of the Sixth International Conference on Teaching Statistics.* Durban, South Africa.

Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 205-237). Boston, MA: Kluwer.

Watson, J. (1997). Assessing statistical thinking using the media. In I. Gal & J. Garfield (Eds.), *Handbook of Assessment in Statistics Education* (pp. 107-121). Amsterdam: IOS Press and the International Statistical Institute.

Watson, J., & Moritz, J. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37* (2), 145-168.

Sample space and the structure of probability combinations in preschoolers.

Z. Nikiforidou, J. Pange
Laboratory of New Technologies and Distance Learning
Department of Early Childhood Education
University of Ioannina-Greece
jpagge@cc.uoi.gr

The sample space and the structure of the possible combinations are components that determine a probability task. The aim of this study is to investigate whether preschoolers show any preference to spatially grouped stimuli and to what extent in terms of number of combinations, they can estimate possible outcomes. There were 3 trials with 2 sub-cases each, with alterations in the position and the number of color-paired sectors. Children showed preference to the paired stimuli and made predictions, based on visual comparisons (Way, 2003,) with ease up to 3 combinations.

Introduction

Risk perception and probability evaluation are concepts that relate to mathematical thinking. Recent studies have shown that preschoolers can make use of the basic probability notions: possible, impossible, sample space (Schlottmann, 2001; Pange & Talbot 2003; Way, 2003, Kafoussi, 2004) by using computational and/or intuitive reasoning skills.

From a traditional theoretical point of view, children are able to concentrate on a single dimension without the ability of reasoning multiple variables (Piaget& Inhelder, 1958). Preschoolers are considered to be static, incapable of causal reasoning and perceptually bound, confined by the similarity of appearance (Piaget, 1965). The ability to relate parts to the whole occurs under computational and operational strategies that emerge later in life. Consistent with this view, children use simplified strategies in probability choice tasks as they cannot take into account alternative possibilities, totality of events, relative frequencies and ratios (Siegler, 1981; Falk & Wilkening, 1998).

On the other hand, some researchers have investigated the intuitive rather than the computational reasoning used by children in random experiments (Shlottmann, 2001; Andreson, 1996; Acredolo et al, 1989). Under this perspective children can relate multiple dimensions and show an intuitive way of counting probabilities (Pange &Talbot, 2003). Children can recognize correlations, make inferences and make use of the frequency of co-occurrence (Kushnir& Gopnik, 2005). Based on visual information, children show a minimal understanding of randomness and can identify the most/least likely outcomes (Way, 2003).

The nature of mathematical thinking in each stage is crucial under the teaching/learning scope. According to Bruner (1960), intuition is a precursor to analytical thinking. Within the classroom, pupils should be encouraged to guess, make hypotheses and predictions, and run the risk of being wrong, before developing their analytical thinking. In addition, children have to be actively engaged while solving a problem in order to acquire knowledge and learn (DeVries et al, 1990; Osborne and Freyberg; 1985).

Based on these educational implications, the current study is going to investigate whether the number of combinations and their position within the sample space has any affect on children's probability estimations. Through their personal involvement in a probability task, participants are tested onto whether their predictions get altered when the sample space undergoes structural and positional changes by maintaining the equality of events.

Methodology

Two groups of preschoolers, in a public kindergarten in Athens in 2006, aged 5 to 6, were tested on to whether they would have a probability understanding of two, three, and four combinations within spatially paired and spatially non paired stimuli. 10 children participated in condition 1 (single-paired condition) and 10 in condition 2 (non-paired condition). For the purpose of the current study, by terms of spatially paired stimuli we mean the allocation of two similar-colored sectors next to each other inside a circle. Any other position of the two similar-colored sectors among the circle is interpreted as spatially non-paired.

All children carried out a pre- test (pre- assessment) in order to investigate whether they possess the concept of sample space or not; whether they realize that the structure of the sample space affects the equality- inequality of likelihood. They were given a circle divided into 4 parts, 3 of which were green and 1 red. They were asked to predict where an arrow was more likely to stop after spinning around (the starting speed of the arrow in all experiments was initialized by the children).

The experiment was conducted as a random game, as "Mr Arrow" (a drawn arrow pinned in the middle of a circle) was very sad and needed their help in order to decide which color he prefers. The experiment was carried out in a meaningful and childish context so that participants would empathize and become willing to assist the hero. Color-combinations were presented on a circle that was divided each time in equal sectors as the number of colors.

In trial A there were 2 combinations: green and red, so the circle was divided into 4 parts,

in trial B, there were 3 combinations: green, red and blue, so the circle was divided into 6 parts

and in trial C, there were 4 combinations: green, red, blue and brown, so the circle was divided into 8 parts.

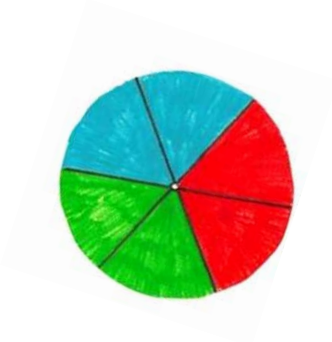
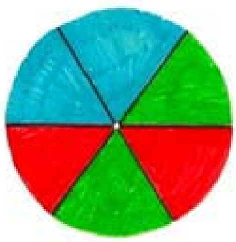
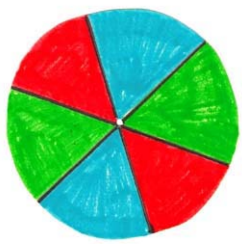
	Condition 1 (single-paired)	Condition 2 (non paired)
1st case (common in both conditions)		
2nd case		

TABLE 1: Example of Trial B (3 combinations: green, red, blue).

Each trial was consisted of 2 sub-cases with a difference in the structure of the combinations in the colors.

Each trial in both Conditions (single-paired and non-paired subsequently) began with the color- combinations presented in ordered pairs (ordered case). The second case of each trial which determined the condition, presented the color- combinations either partly non-paired (Condition 1, where only one single color kept its pair close to itself) or totally non-paired (Condition 2, where all the color- pairs were mixed up).

Children (N=20) ran the pre- test and the three main trials (trial A, trial B and trial C), in pairs in their classroom. Participants had the possibility to interact with each other and in turn they were asked by the researcher to predict which color "Mr Arrow" would show. Then, they were asked to push him with the finger and spin him around in order to find out his choice. Responses were tape-recorded and used for further analysis.

Results

Children were quite enthusiastic to help Mr Arrow. At this age, children empathize and become identical with the heroes of the stories. The meaningful context allowed them to get actively engaged:

Researcher: …so, what do you think, are you going to help him?

Tasos: Yes, he seems nice.

Amalia: I don't want him to be sad…

In the pre-test trial 15 out of the 20 children gave the 'expected' answer; they could predict that green was the most likely color to come out, as it possessed the ¾ of the sample space:

Researcher: Which color do you think Mr Arrow will show?

Andreas: He will show green, green is more.

Maria: Green. There are 3 green.

The rest of the children, answered intuitively, based on random justifications:

Researcher: Why red?

Tasos: I like red, so red will appear.

Niki: Mr Arrow is a boy so he will prefer red.

A. In the first trial we used two colors (i.e. red, green) and children made predictions by chance. Others predicted red and others green (equality of likelihood).

Based on the outcome of the 1st case, 13 children in total predicted the same color as a possible outcome in case 2, no matter if the sample space was ordered or unordered. This implies that children of this age have an understanding of co- occurrence:

Popi: He will show red, as before…

B. With the addition of the third color (i.e. red, green, blue) in the second trial, blue (the new color) was the most frequent prediction among the responses of the ordered case.

In the single paired case of this trial (Condition 1), 7 children predicted the color that was still ordered and spatially next to its pair:

Researcher: What about this circle? What do you think Mr Arrow will show now?

Nikos: Blue, there are more blue triangles.

Dimitris: He will prefer the blue, because it is all together.

The alteration of the colors' positions had an affect on children's predictions. The fact that blue would keep an ordered arrangement, opposed to the rest colors, made it the most possible outcome.

In the non-paired case, 6 children again based their answers on the outcome of the last case. It is worth mentioning that all of them had realized that the positions of the colored parts had changed and that no new color was added:

Researcher: Let's see this circle! Which color do you think Tina that Mr Arrow will show this time?

Tina: Look, the colors are the same…

Lena: But they are up and down…

Tina: Never mind… he showed green before… I think he will show green. (response based on co-occurrence)

Children in this case were able to follow the changes in the combinations. They could visually recognize that the positions of the 3 combinations had been mixed up. They understood that no new color was added in the circle and that red, green, and blue combinations had just been presented in different positions. They were in the position of making visual comparisons. Again, previous outcomes affected children's predictions for the next trial.

C. In the 1st case of the third trial, 9 children in total predicted the new color. The new visual input again (as in trial B) influenced their predictions:

Aggelos: Brown is new! Mr Arrow will prefer it…

However, children seemed not to be able to follow all 4 combinations (i.e. red, green, blue and brown). They started to loose their concentration and paid attention to other factors, for instance:

Researcher: Let's move on to this circle… [interruption]

Liana: Again these colors? When will pink appear?

Mina: Will Mr Arrow visit us again?

Researcher: Sure, he may also bring Mrs Arrow. But, come on, let's finish. Liana it is your turn, which color do you think he will choose?

Liana: Brown.

Researcher: Why brown?

Liana: Because brown (intuitive answer).

Research: Ok then, let's see.

In the single-paired case (condition 1), this time only 5 children predicted the 'ordered' color as a possible outcome.

In condition 2 no pattern was found in the children's predictions. In general the responses were based on chance:

Efi: Blue was before, now it will be green…

Researcher: Why do you think green will appear?

Efi: I don't know.

Children seemed to be tired at this point. Answers had no justification and children had in a way lost their interest.

Discussion

The most interesting results relate to the structure of the sample space. The position of the stimuli affected the children's responses. Children favored the spatially paired combinations. This is quite evident in the second cases of the 1st Condition, where children preferred the color combination that was presented next to its pair. In the 2nd Condition where colors were not allocated as subsequent pairs on the circle, children responded randomly or intuitively.

Of course, the fact that the design of the current experiment was based on paired stimuli in this precise way may imply that children may have worked on other meanings too, such as proportions or quantified categorization rather than on mere probabilities (Deák& Bauer, 1996); this is an issue for further investigation. Thus, the focus of the present study was to find out whether preschoolers could understand the equality of events when the same color combinations were visually presented in different positions among the circle.

These results comply with the theory that at this age, children respond intuitively (Schlottmann, 2001; Pange & Talbot, 2003) based on visual perception, abstract reasoning and personal preferences. Children showed the ability to focus on more than one variable (color and position) in contrast to the piagetian notion of centration. In addition, it was evident in the first trial of both Conditions that children were able to follow and estimate the frequency of co-occurrence. In accordance with the findings of previous studies, children were able to make inferences and predictions based on what occurred previously. Evidence has shown that even at the age of 3, children have the ability of inference making and similarity selecting (Deak et al, 2002; Kushnir& Gopnik 2005).

Children showed a strong reliance on visual impressions and comparisons; they could see and understand that the structure of the sample space had changed but they couldn't recognize that there was still equality of likelihood. They were also affected by the entrance of a new color and in the beginning of each trial they would use this color as the color of their prediction. Preschoolers were able to notice that the same number of paired combinations appeared in different positions in both cases but they couldn't estimate the same possibility for each color. In consequence, the structure of sample space and the likelihood of events do not have a clear connection at this age (non-probabilistic thinking stage; Way 2003).

Finally, the above results imply that preschoolers can be easily aware of 3 combinations and begin to have difficulties in their predictions when a fourth element enters the task. At this point, children begun to loose the track of the task and responded by chance. For instance, 3 girls complained about the 4th color (brown) as they would rather expect a more 'female' one. Such findings relate to other studies that have shown that infants can represent the exact number of objects in a scene, up to a set size limit of three or four (Barth et al, 2005).

The fact that the current experiment took place in a meaningful context as a 'game- mystery', instead of a context with random shots and no sequence and story, enabled the children to be actively involved in the task through their new friend. Participants were personally engaged in order to help Mr. Arrow and their own intentions were identified with those of Mr- Arrow and vice versa. Such approaches where an external figure or puppet, ie Mr Arrow, are part of the experiment are quite often in studies with children of this age (Denison et al, 2006; Cimpian& Markman, 2003; Doherty, 2002; Siegler, 1991), so that they get motivated within a context. Under these guidelines, children were encouraged to predict, estimate, and construct knowledge (basic concepts of the constructivist approach; DeVries et al, 1990; Osborne and Freyberg; 1985). Thus, this sort of interference might have influenced children's pure stochastic way of thinking.

These results may be useful under the perspective of instruction-teaching, as science content and cognitive capacity are important to match (Lind, 1999). Preschoolers can manipulate pairs of 3 combinations but when it comes to the fourth they find it difficult. Moreover, further research is needed to investigate more profoundly how intuitive and computational reasoning develop and whether there is interplay, as well as how intuition is complementary to formal understanding (Bruner, 1960).

Further research could investigate how the different combinations of a sample space (up to three) are processed through other probability task designs (with methodological alterations). For instance, color combinations can be replaced by items or other stimuli (i.e. pictures of animals or tools); paired combinations can be replaced by single ones in order to eliminate the possibility of children's working with mechanisms of categorization or identical observation; the sample space can be represented in an ordinal or vertical rather than a circular space. This sort of visual modifications may affect the children's responses and may imply other results i.e. extension of 3 combinations to 4; this needs replication. In addition, computer – based tasks may rouse new findings too. Here, we entail the issue of the utility of New Technologies in the classroom?

To sum up, such methodological and theoretical implications can be investigated more profoundly and as a result be incorporated in the teaching- learning procedure of probabilities in preschool education. So, if we would like to move on to a more formal learning procedure like the construction of knowledge, then the same task should be repeated in other contexts (i.e. computer- based tasks), under precise educational goals.

REFERENCES

Barth H.,La Mont K., Lipton J., Spekle E. (2005). Abstract number and arithmetic in preschool children. *PNAS* 39 (102): 14116- 14121.

Bruner J. (1960). *The Process of Education*. Cambridge, Mass.: Harvard University Press.

Deák, G. & Bauer, P. J. (1996). The dynamics of preschoolers' categorization choices. *Child Development*. 67, 740-767.

Denison S., Konopczynski K., Garcia V., Xu F. (2006). Probabilistic Reasoning in Preschoolers:Random Sampling and Base Rate. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 1216-1221.

DeVries and Kohlberg (1987/1990). *Constructivist early education: Overview and comparison with other programs.* Washington, DC: National Association for the Education of Young Children.

Doherty M. (2002). Synonyms, homonyms and false beliefs: a reply to Garnham,Brooks, Garnham and Ostenfeld. *Developmental Science 5:*1 (2002), pp 84–85

Kamii and DeVries (1978). *Physical knowledge in preschool education: Implications of Piaget's theory.* Englewood Cliffs, NJ: Prentice Hall.

Kushnir T. & Gopnik A. (2005). Young Children Infer Causal Stenght From Probabilities and Interventions. *Psychological Science* 16 (9): 678-683.

Lind K. (1999). Science in Early Childhood: Developing and Acquiring Fundamental Concepts and Skills. *American Association for the Advancement of Science*.

Osborne M. and Feyberg P. (1985). *Learning in science: Implications of children's science.* Aucland, New Zealand: Heinemann.

Piaget J. & Inhelder B. (1975). *The origin of the idea of chance in children.* London: Routledge & Kegan Paul. (Original work published 1951).

Pange J., Talbot M. (2003). Literature Survey and Children and their perception of Risk. *ZDM* 35(4): 182-186.

Pange J. (2003). Teaching Probabilities and Statistics to Preschool Children. *Information Technology in Childhood Education Annual* (1):163-172.

Kafoussi S. (2004). Can kindergarten children be successfully involved in probabilistic tasks? *Statistics Education Research Journal* 3 (1), 29- 39.

Siegler R. (1991).In Young Children's Counting, Procedures Precedes Principles.*Educational Psychology Review, 1Iol. 3, No. 2, 1991*

Schlottmann A. (2001). Children's Probability Intuitions: Understanding the Expexted Value of Complex Gambles. *Child development* 72 (1): 103- 122.

Way J. (2003). The development of young children's notions of probability. *Proceedings of CERME*. Bellaria. Italy.

http://www.stanford.edu/~acimpian/CDS_2003_poster.pdf

LOOKING FOR RANDOMNESS IN TASKS OF PROSPECTIVE TEACHERS

Efi Paparistodemou*, Despina Potari** and Demetra Pitta***

* Ministry of Education, Cyprus

**University of Patras, Greece

***University of Cyprus, Cyprus

This study investigates the way that prospective teachers conceive the role of tasks in pre-primary mathematics teaching regarding randomness. This is explored in the context of the teaching practice of prospective teachers during the last semester of their university studies. The methodology of the study was organised in four stages: 1. design of lesson plan; 2. classroom implementation; 3. discussion of the lesson with the school practice instructor and; 4. self-assessment report and redesign of the lesson. These stages have been acknowledged as crucial settings for encouraging reflection, awareness and professional development in general (Moyer,2001). In the particular study we analysed four cases of prospective teachers who planned and taught lessons regarding probability. The data comes from their lesson plans, observations of their teaching, interviews after their teaching and their written self-assessment reports. The analysis of the data shows that prospective teachers appreciated the importance of motivation and using tools in their classrooms for teachings stochastics. However, from our classroom observations we identified that the activity was often mathematically trivialized and that there was a misunderstanding about the stochastic interpretation of the task.

INTRODUCTION

Hoyles (2002) claims that the mathematical activity is designed to foster mathematical meanings through construction, interaction and feedback, and also the students could scaffold their own thinking through communicating with the tools. However, these tools are often considered early years more as means to motivate children rather than to challenge them mathematically. One reason could be teachers' lack of sound mathematical knowledge that would have allowed them plan activities that would enhance young children's mathematical activity. The importance of a strong knowledge of content in teaching mathematics has been acknowledged by a variety of references (e.g. Mason, 1998; Ma, 1999). In addition to this, teachers need to consider children's ideas and intuitions both in their planning and in their actual teaching. Shulman (1986) defines this knowledge as 'pedagogical content knowledge' and Ball and Bass (2000) believe that teaching mathematics entails work with microscopic elements of mathematical knowledge in order to make sense of a child's apparent error or appreciate a child's insight.

The integration of mathematics and pedagogy is described by Boaler (2003) as "mathematical know-how" and it can be achieved through teachers' reflection on their own teaching. Prospective teachers operate from a complex knowledge base, both pedagogical and mathematical, which is developed through their personal experiences tacit and academic. Mason (1998) also appreciates the importance of reflection. According to him the key notions underlying real teaching are the structure of attention and the nature of awareness. During their teaching practice as prospective teachers design lesson plans, teach and evaluate their teaching the question of what they are "noticing" in these phases is an essential one. Moreover, their own interpretations of the things that they notice give an indication of their level of awareness. In the case of mathematics teaching, the meaning we attribute to awareness in action is related to the craft knowledge that Ruthven (2002) describes. It is related to the teaching tools - materials that the prospective teacher brings to the classroom either from her own experience as learner of mathematics or either from her experience at the university in the area of mathematics education. In that level the arguments that she develops to support her teaching decisions and choices are grounded in this craft knowledge and remain at a primitive level. On the other hand, the awareness in the discipline of mathematics education, in our view, is the result of relating the academic knowledge, which is a result of university experience, but it is integrated to the craft knowledge that the prospective teacher possesses.

Nowadays, probability and statistics have an important role to play in our everyday life, especially at children, where most of their games have the idea of chance. It is natural for humans to use statements of probability to describe uncertainty of the external world. People speak in everyday terms of 'chance' and 'randomness'. These concepts often serve them well in everyday communication because of the general consensus about their meanings. Yet, randomness is one of the most elusive concepts in mathematics. Hacking (1975) suggested that the meaning of the word random could be answered briefly, but it would take 100 pages to prove any answer correct! Moore (1990) implies an external source of the uncertainty, referring to the external world, and opposed to an internal source in the form of one's knowledge. Falk, Falk, and Levin (1980) argue that probability is composed of two sub concepts: chance and proportion. One has to be aware of the uncertain nature of a situation in order to apply the results of proportional computations. Obviously the ability to calculate proportions as such does not necessarily signify understanding of probability. A realisation of uncertainty either in controlling or in predicting the outcome of an event is crucial.

Paparistodemou and Philippou (2002) describe how young children start to make probabilistic decisions and think about chance and risk from an early age, depending on *how* they have embarked on probabilistic games. According to Borovcnik and Peard

(1996), there is no doubt that the topic of probability is an important one in the mathematics curriculum even though the inclusion of probability is a relatively recent development. Research (Paparistodemou, 2004; Paparistodemou & Noss, 2004; Pratt, 2000) also shows that the design of an activity at the age of 4-11 is very important for children to express probabilistic ideas. According to Borovcnik and Peard (1996), there is no doubt that the topic of probability is an important one in the mathematics curriculum even though the inclusion of probability is a relatively recent development.

Petocz and Reid (2002) indicate the importance for development of learning environments that can engage students' interest, broaden their understanding of statistics and enrich their own lives. The complexity of mathematics teaching requires the development of awareness in a number of different elements that constitute teaching. Jaworski (1994) has identified some of these aspects: management of learning, mathematical challenge and sensitivity to students. These three are integrated under the notion of teaching triad. In this study, based on the Teaching Triad model (Potari & Jaworski, 2002) and Mason's ideas on reflection and awareness (Mason, 1998), we investigate the kind of mathematical challenge prospective teachers offer to their students concerning the concept of randomness.

METHODOLOGY

The methodology of the research follows a qualitative approach. The participants are case studies of four prospective pre-primary teachers who were doing their teaching practice in pre-primary schools as a part of their university degree. The participants have passed successfully the following courses, which concerned stochastic and teaching mathematics, before participating at their teaching practice: *Statistical Methods*: descriptive statistics, probability, binomial and normal distribution, sampling, confidence intervals, hypothesis testing, correlation, regression analysis, introduction to analysis of variance; *Pre-Math Concepts:* basic theoretical trends in Psychology concerning the development of pre-mathematical concepts in early childhood, the importance of language in the development of the first mathematical concepts, critical analysis of the arithmetic of natural numbers*; Mathematical Concepts in the Kindergarten school*: the content of mathematics for the kindergarten and the first grades of the primary school, the teaching methods of the subject as they have developed in recent years, the teaching aids, and the contemporary methods of evaluating the mathematical ability of pupils, fundamental psychological theories as they concern the development of primary mathematical concepts in preprimary school children.

The data was collected from the four following sources: Source 1- Lesson plans: Each prospective teacher prepared four lesson plans on stochastics to the same pre-primary classroom (age of children 4-5.5). The lesson plans were given to the first author before their teaching lesson. Source 2 – Observations of their teaching: The first author observed the prospective teachers lessons and took notes during their teaching. The role of the researcher was that of a marginal participant (Robson, 1993). Source 3 – Interviews after their teaching: The participants were interviewed after their taught lessons. The participants expressed their first reaction on their lesson in a 'semi-structured' interview (Scott and Usher, 1999). Source 4 - Written self-assessment reports: Prospective teachers wrote self-assessment reports after their teaching, where they expressed their thoughts concerning the positive and the negative aspects of their teaching. In the self-assessment reports teachers had the chance to modify their lesson plan.

The data was analysed by identifying extracts of related to mathematical awareness. Mathematical awareness was approached through the reflection process of the prospective teachers, during the interview and the self-assessment report. In particular, a number of critical events (like connecting mathematical activities to the task) were identified by the researchers from the lesson plan and the classroom observation. The critical events were not mentioned from the beginning of the interview. The prospective teachers either commented on these spontaneously or not. In the latter, they were prompted by the researchers to discuss about them. The researchers produced a teacher profile for each participant from all four sources. A validity check between the researchers was made when a researcher identified a critical event and a teacher profile was produced at the end. The data in this study is analysed concerning mathematical challenge and the concept of randomness and in the following paragraphs a characteristic example of the data is presented.

MATHEMATICAL CHALLENGE AND THE CONCEPT OF RANDOMNESS: PROSPECTIVE TEACHERS' MATHEMATICAL AWARENESSS

Logical versus stochastic interpretations of possibility

According to their academic background, prospective teachers were trained in statistical methods and mathematical concepts in kindergarten school. It is interesting that although the focus of prospective teachers' lesson was stochastic interpretation, many of their tasks did not include the concept of randomness. From the analysis of the data, we realized that there was confusion between logical and stochastic interpretation of possibility. In the logical interpretation of possibility the concept of randomness and chance is missing. For example, a prospective teacher, Amy[1], planned a seven-activity lesson. One of her tasks shows an example of what we consider as a logical interpretation of possibility: 'from a non-transparent bag find the possibility of choosing a soft ball or a book after you place your hand inside'. Here, children did not make any stochastic interpretations as when they place their hand inside they recognized which one was the soft item. In this tasks there is absence of randomness.

[1] All the names are pseudonyms

Amy's first activity concerned some pictures with a scared duckling in a forest. She asked children to *guess* what happened in a forest and resulted to the duckling being scared. She explained in her lesson plan that the aim of this activity was for children to understand randomness by using the words 'may be', 'perhaps', 'possible', 'impossible'. In her second activity she mixed up pictures of animal mothers and their children. She asked children to match the animal mothers of animal children that had been lost in the forest. She explained that the aim of this activity was also children to use the words 'possible', 'impossible' and 'certain'. She stated in her interview:

> 'Children liked these activities and they used the new words! For example, because the duckling was in a forest it would be *impossible* to be scared from a car'.

In these two activities we can see that children used the words 'impossible', 'may be', 'possible'. We agree that here the logical use of 'possibility' can be a reasonable precursor to later work on the stochastic treatment. But, since randomness is missing from the task there is not a stochastic interpretation. Students were asked to answer whether something was possible, certain or impossible without being presented with a probabilistic scenario. The prospective teacher appeared to be using these words with children, but we can say that children have not been directed towards stochastic interpretation. In another activity children asked to make a guess 'of what may be the drawing hidden behind a carbon by partially revealing it'. Her aim was to express the word possible. However, students were not directed towards the mathematical content of probability, since the element of uncertainty and randomness was missing. Thus, although children might reason what the picture might be, they did not really approach the idea of probability. In the interview the researcher asked her to explain how she understood the probability (R stands for the researcher):

> R: What does the concept of probability mean to you?
>
> A: The may be…that something is possible.
>
> R: What does it mean when we say 'something is possible'?
>
> A: …I know the word…it is at the edge of my tongue…but it doesn't come to me now…Well, not to know something…

We can argue here that the idea of probability was confused to 'what something may be' instead of 'what may happens'. Amy appeared not to realise that the probability is related to the likelihood of an event occurring and she related it with the likelihood of something being a certain object. In Amy's words we identified some 'apprehension' in her words. We could argue that the teacher based her definition to her own intuitions and she has not been able to identify the mathematical meaning of what a probability is.

In Cathy's case we can also recognise the concept of 'randomness' in the aims of her lesson. Again, children used the words 'possible', 'impossible, 'certain', with the absence of randomness. For example, the prospective teacher said that she had a number of pictures in her hands and children had to decide which of the pictures could have been taken in a park. The children had to concentrate on some aspects of the picture, i.e. darkness, and decide whether these could have been seen in a park. Children could answer that it was certain, possible or impossible. The teacher tended to introduce imaginative situations in which more than one eventuality might apply but that scenario can be confused with one which is really about probability. We believe that the probabilistic words certain, impossible, possible, uncertain needed to be discussed in the classroom however the lesson (as in this case) should not be limited to this. In her interview she argued:

> "Is this connected with guessing? Would it be more ok if I had a bag and draw objects from inside? … I mean to predict."

In her words we can identify a kind of questioning of what is randomness. Cathy in her self-assessment report stated that she felt happy about her lesson and explained that her lesson went well because the children were excited, although some teachers consider a probability lesson to be difficult. It can be argued that she took in consideration children's reaction to the activity. However, we feel that she considered whether they enjoy the activity but not the learning outcomes. Of course this is not surprising since we feel that she did not know what the learning outcomes should have been and therefore she could not look for them in children's responses. Moreover, Cathy explained that a limitation she found was that she did not have the pictures ready on the board and the fact that she had to place them during the lesson was time consuming.

Selecting a spinner

In Macy's lesson plan we recognised an awareness of the concept of randomness. Macy was very specific in her two aims of the lesson, which were 'Children should be in a position to select the appropriate sample spaces for certain, impossible and fair events' and 'Children should experience the concept of randomness'. In her first activity she was planning to show children 6 spinners, which had different sample spaces (for example, whole blue, whole red, half blue half red, ¾ blue and ¼ red). Children would be asked questions such as:

> 'Which spinner should the blue boy use in order to win? Why?
>
> If I spin this spinner will I get blue for sure? Why?'

Macy seemed to be very aware of the epistemological characteristics of her activity in her lesson plan and her task seemed well organized for introducing a stochastic interpretation. Although in the planning of the activity she stated that children would experience the concept of randomness, when she was teaching she presented to the children the spinners and asked them what colour would get without spinning any of the spinners. In addition to this, her questions were often phrased to direct specific answers. For example, 'For blue to win shall we get the blue spinner?' This had as a result children providing yes or no answers and not using the words certain, impossible, possible. Macy's case is interesting from the point of view that although we recognised from her lesson plan a mathematical challenge awareness, we can see from her teaching that she was not aware of a deeper understanding of the probability concept and the presence of randomness in her activities. On this point, in her interview she stated:

M: This activity aimed at children seeing different sample spaces and making decisions in regard to which colour would win.

R: What did you want children to learn?

M: …Different sample spaces…Well, the activity went well, so the level of difficulty was good and children did not develop any misconceptions.

We can recognise here that Macy used words like 'sample spaces', 'making decisions' and reflects on her activity on mathematical issues. Besides that, we can say that there is a limitation on this mathematical challenge, as she did not recognise the importance of randomness on her probabilistic activities. It is important to mention here, that Macy in her self-assessment report refers to the importance of randomness in her tasks and make changes to her lesson plan according to this.

Mathematical limitations

In their self-assessment reports some teachers seem to realise some mathematical limitations. Evelyn expressed her thoughts as follows:

> 'I feel that the lesson did not go so well for several reasons. Some of them are: the fact that I had to think about the activities by myself [she argued earlier in her interview that she couldn't easily find activities in bibliography for this theme], the fact that I was not sure of the concept of probability myself, led to the lesson plan not being very good. However, I feel that the lesson would have been better if the children were calm'.

Evelyn was aware that if she had a deeper understanding of the probability concept this would have helped her to improve her planning. Besides that, she also reflected on management learning to address what she would have changed on her lesson, after her teaching.

DISCUSSION

Quite often the mathematical challenge seems to be rather trivial both in the planning and in the classroom teaching (Jaworski, 1994). Although they state the mathematical aims of the lesson, in many cases these aims appeared to have certain limitations. For example, sometimes they were very general, emphasized procedures or disconnected from the designed tasks. One explanation for this appeared to be prospective teachers' lack of mathematical awareness. The data has shown that although the aims of the stochastic lesson plans were for children to use and understand the words "certain", "probable", "impossible" in a random scenario, children were using verbally these words in various scenarios without using the idea of randomness. The verbal use of these words might be a reasonable precursor to later work on the stochastic treatment. But, since the aim of the lesson plans was refer also to the concept of randomness, the existence of the concept in the tasks is really important (see Falk, Falk, and Levin,1980)

Prospective teachers were concerned mostly of affective aspects and whether children were engrossed in the activities. For example, they were interested whether children had "fun". However, prospective teachers did not appear to pay much attention to students' intuitive ideas, prior knowledge, queries or quality of children's responses and thinking. Even in cases where they focused on the above issues and showed some elements of awareness in action, they were not able to interpret these and move to a higher level of awareness in discipline. As a result, a balance between the affective and the cognitive side of sensitivity in many cases was not achieved.

The management of learning was the domain that seemed to attract most of their attention (Mason, 1998). This is not surprising if one considers that probably one of the main concerns of a prospective teacher during their teaching practice is to "survive" in the classroom. Thus, group work, use of manipulatives, games, time management, were incorporated in their teaching. Although these approaches show a degree of awareness in action, their potential value was not achieved since these pedagogical practices were not interwoven successfully with the mathematical ideas. This could only be achieved if the prospective teachers had awareness in the discipline of mathematics education

The data has shown the kind of experiences that guide prospective teachers' design, implementation and evaluation of teaching. Prospective teachers appreciated the importance of using tools in their classrooms for teaching stochastics. However, these tools are often considered more as means to motivate children rather than to challenge them mathematically (Hoyles, 2002). It can be argued that prospective teachers build relations between theory and practice at a rather general pedagogical level. The transition and reflection to more specific mathematical and pedagogical issues appears to be important to the 'mathematical know-how' procedure (Boaler, 2003), but a difficult endeavour. This calls upon special attention and reflection on behalf of mathematics teacher educators to tackle this problem.

The training that the teachers had undergone concerned statistical methods and mathematical concepts in kindergarten school. Prospective teachers operate from a complex knowledge base, both pedagogical and mathematical, which is developed through their personal experiences tacit and academic. The study shows that there is a mismatch between the knowledge being constructed in that training and how that knowledge can be used to teach young students. An implication of the study is that the teaching practice of these teachers could include a procedure, like the methodology that the this study adopted, were the prospective teachers can reflect and get feedback on their mathematical awareness according to their teaching experience. The stages of the methodology have been acknowledged as crucial settings for encouraging reflection, awareness and professional development in general (Moyer,2001). Mason (1998) appreciates the importance of reflection. In our view awareness in the discipline of mathematics education is the result of relating the academic knowledge, which is a result of university experience, but it is integrated to the craft knowledge that the prospective teacher possesses.

REFERENCES

Ball, D. L. & Bass, H. (2000). Interweaving content an pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on the teaching and learning of mathematics* (pp. 83-104). Westport, CT: Ablex.

Boaler, J. (2003). Exploring the nature of mathematical activity: using theory, research and 'working hypotheses' to broaden conceptions of mathematics knowing. *Educational Studies in Mathematics, 51,* 3-21.

Borovcnik, M., and Peard, R. (1996). Probability in Bishop, A., Clements, K., Keitel, C. Kilpatrick, J. and Laborde, C. (eds.) *International Handbook of Mathematics Education* Netherlands: Kluwer Academic Publishers, p. 239-287.

Falk, R., Falk, R., and Levin, I. (1980). 'A potential for learning probability in young children' Educational Studies in Mathematics, 11, p. 181-204

Hacking, I. (1975). The emergence of probability Cambridge: Cambridge University Press

Hoyles, C. (2002). From describing to designing mathematical activity: the next step in developing a social approach to research in mathematics education? *Educational Studies in Mathematics*, *46,* 273-286.

Jaworski, B. (1994). *Investigating mathematics teaching: A constructivist enquiry.* London: Falmer Press.

Ma, L. (1999). *Knowing and teaching elementary mathematics.* Mahwah, NJ: Erlbaum.

Mason, J. (1998). Enabling teachers to be real teachers: necessary levels of awareness and structure of attention. *Journal of Mathematics Teachers Education, 1,* p. 243-267.

Moore, D. (1990). 'Uncertainty' in Steen, L. A. (ed.) On the shoulders of giants: New approaches to numeracy Washington: National Academy Press, p. 95-137

Moyer (2001). Are we having fun yet? How teachers use manipulatives to teach mathematics. *Educational Studies in Mathematics*, *47,* p. 175-197.

Paparistodemou, E. (2004). *Children's Expressions of Randomness: Constructing Probabilistic Ideas In an Open Computer Game,* Institute of Education, University of London (PhD Thesis).

Paparistodemou, E. and Noss, R. (2004). Designing for Local and Global Meanings of Randomness' continuum *Proceedings of the Twentieth Eighth Annual Conference of the International Group for the Psychology of Mathematics Education Vol. 3,* p. 497-504, Bergen, Norway.

Paparistodemou, E. and Philippou, G. (2002). 'Intuitions of the Concept of Probability in 6-7 year-old Children' *Themes in Education, 3*, 1, p. 63-78.

Petocz, P. and Reid, A. (2002). How students experience learning statistics and teaching *Proceedings of the 6th International Conference of Teaching Statistics (ICOTS 6)* Cape Town.

Potari, D. & Jaworski, B. (2002). Tackling Complexity in Mathematics Teaching Development: using the teaching triad as a tool for reflection and analysis. *Journal of Mathematics Teacher Education, 5*, p. 351-380.

Pratt, D. (2000). Making sense of the Total of Two Dice *Journal for Research in Mathematics Education, 31*, 5, p. 602-625.

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

Ruthven, K. (2002) Linking researching with teaching: towards synergy of scholarly and craft knowledge. In L. English (Ed.) *Handbook of International Research in Mathematics Education* (Lawrence Erlbaum, Mahwah NJ) p. 581-598.

Scott, D. and Usher, R. (1999). *Researching Education* London: Cassell.

Shulman, D. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15, p. 4-14.

MAKING CONNECTIONS BETWEEN THE TWO PERSPECTIVES ON DISTRIBUTION

Theodosia Prodromou t.prodromou@warwick.ac.uk

Centre for New Technologies Research in Education, University of Warwick, UK

My premise, in line with a constructivist approach and Pratt's (1998) research, is that thinking about distribution must develop from causal meanings already established. The results of the third iteration of a design research study indicate support for my conjecture that it is possible to design an environment in which students' well established causal meanings can be exploited to coordinate the emergent data-centric and modelling perspectives on distribution (Prodromou & Pratt, 2006). In this study, I report on the fourth iteration that investigates how and whether students bridge the two perspectives on distribution.

TWO PERSPECTIVES ON DISTRIBUTION

Prodromou & Pratt (2006) referred to the *modelling* and *data-centric* perspectives on distribution which offer different views of variation. In the *data-centric* perspective, data is seen as spread across a range of values. In the *modelling* perspective, movements are considered as random under some probabilistic mechanism away from the main effect. A *data-centric* perspective on distribution which is in fact an empirical distribution pays attention to the variation and shape of data that has been collected, perhaps through a sampling process. Thus an empirical distribution is one composed of a set of variables that has been collected or is capable in principle of being observed.

Exploratory Data Analysis (EDA) approach promotes a perspective on distribution as a representation of collections of actual data. The introduction of ICT into schools has prompted interest in Exploratory Data Analysis (EDA) as a means of engaging students in statistical analysis, arguably reducing the need for a sophisticated understanding of probability theory prior to meaningful engagement. Dynamic visual displays (Biehler, 1989) are ideally suited to support students as they manipulate data and use a range of different representations in order to infer underlying trends.

There are many research studies that take a *data-centric* perspective on distribution. Previous research has conceived of distribution as "an important part of learning to look at the data" (Moore, 1990, p.106) and as an organising conceptual structure with which we can observe the aggregate features of data sets rather than just individual values (Cobb, 1999). Perusal of recent research literature suggests that reasoning about variation and distributions are strongly associated (Ben-Zvi, 2004).

When we refer to the *modelling* perspective, we refer to the theoretical distributions (for example, Normal, Uniform and Binomial) that are mathematical models, in which we attribute probabilities to a range of possible outcomes (discrete or continuous) in the sample space. In this modelling approach a mathematical model, in which we attribute probabilities gives rise to variation. Data distributions are regarded as variations from the ideal model, the variations being the result of noise or error randomly affecting the signal or the main effect, as reflected in the model it self. The signal, therefore, can be an average value with variation as noise around it (Konold & Pollatsek, 2004) or a distribution, such as the shape of a smooth bell curve of the normal distribution, with which we model data (Bakker, 2004).

The *modelling* perspective of distribution pays attention to randomness and the probabilities that mould a wide range of scientifically real-world phenomena or the outcomes of some experiment. The *modelling* perspective reflects the mindset of statisticians when applying classical statistical inference.

CONNECTING THE TWO PERSPECTIVES ON DISTRIBUTION

Piaget and Inhelder (1975, translated from original in 1951) studied the idea of stochastic convergence as a function of the mental development of the subjects. Only at the third stage do 11- to 12-years-olds, according to Piaget, appreciate the role of large numbers in the regularity of a distribution which, in turn, allows for some beginnings of gradations up to the discovery of the normal distribution. In this sense, Piaget offers us the first hint that we only begin to gain some mastery over the normal distribution when we are able to grasp the role of large numbers in the regularity of distribution.

Prodromou and Pratt (2006) claimed that the concept image of distribution might be seen as impoverished if it were not to recognise its role as a probabilistic model for various types of phenomena as well as incorporating the *data-centric* perspective. They suggested that the concept image of distribution lies in coordinating emergent *data-centric* and *modelling* perspectives on distribution. They also argued that the emphasis of contemporary curricula on EDA approaches is insufficient alone to nurture such co-ordination. In order to coordinate these two perspectives, Prodromou and Pratt (2006) argued that it is necessary to see them as a duality that encompasses both the deterministic and the stochastic.

Piaget & Inhelder (1975, translated from original in 1951) suggested that the learner fails in the first place to apply operational thinking to the task of constructing meanings for random phenomena. Only much later, according to Piaget, the learner succeeds in inventing probability as a means of operationalising randomness. Piaget offers a first hint that we only begin to gain mastery over the stochastic when we know how to exploit our well-established appreciation of the deterministic. Piaget's constructivist stance demands that we take into consideration the prior knowledge of students since therein must lie the resources for appreciating distribution and other core stochastic concepts.

Pratt (1998) reported that the *local resources* brought to bear by students of age 11 in describing short-term randomness, were remarkably akin to those of experts in one respect. Nevertheless, the same students were unable to demonstrate meanings for distribution or the *Law of Large Numbers. Global resources* however began to emerge as these students engaged with specially designed tools, in a microworld called *ChanceMaker*. The students began to articulate new meanings for the longer term. These meanings were causally-based and situated versions of the Law of Large Numbers, such as "the more trials you do, the more even is the pie chart". By phenomenalising (Pratt, 1998) randomness, Pratt claimed that the students were able to exploit well-established knowledge about causality to concretise (after Wilensky, 1991) the *Law of Large Numbers*.

Pratt's work is significant for the present study due to the causal nature of students' resources. It makes a prima facie case that technologically-based environments may have the potential to facilitate the construction of *global resources* out of causality. Prodromou and Pratt (2006) conjectured that, given appropriate phenomenalised tools, students would be able to coordinate the *modelling* and *data-centric* perspectives of distribution. They designed a BasketBall microworld in which randomness could become an agent that causes variation and in turn randomness could be "controlled" through parameters instantiated as on-screen sliders; from an expert point of view they might be perceived as measures of average and spread. The handle on the slider and the arrows are initially seen as ways of controlling the throwing of the basketball. Later however they become a representation of the act of throwing (see Prodromou & Pratt, 2006). The mechanism for this fusion between a measure or representation and a control is what Papert (1996) called the *Power Principle;* students coming to know through use. Prodromou and Pratt (2006) noticed, however, that there was a paradox. On the one hand, Pratt (1998) makes a prima facie case that technologically-based environments may have the potential to promote a method of constructing meanings for distribution out of causality. On the other hand, such an approach may strengthen the "centralised mindset" (Resnick, 1991) and militate against the construction of distribution as an emergent phenomenon (Prodromou, 2004) that bridges the *data-centric* and *modelling* perspectives on distribution.

This study is the continuation of the work of Prodromou and Pratt (2006). Both studies are based on my doctoral research. To elaborate the research question of this study, I aimed first to instantiate the conjectures into a new version of Basketball microworld that would perturb the students' thinking and act as a window on that thinking-in-change (Noss & Hoyles, 1996) about the two perspectives on distribution.

METHOD

Influenced by the Constructionists' (Harel & Papert, 1991) accent on the affective, I placed emphasis on developing activities for a playful context in which students are likely to construct *purpose*, while at the same time coming to appreciate the *utility* (Ainley, Pratt & Hansen, 2006) of distribution as a central concept. The current study falls into the category of design experiments (Cobb et all, 2003) which, in combination with the delicate process of *phenomenalising* a mathematical concept that can capture learners' needs by transforming powerful ideas into situated, meaningful and manipulable phenomena, gradually sensitise us towards the complex learning ecology of the domain being investigated. Typically designed experiments require successive iterations to be used by students often weeks apart. The design is informed by analysis of the students' activities in the previous iteration. In this paper, I report on students' interaction with the fourth iteration of the microworld. In the third iteration, I conjectured that I would be able to build an environment that enabled the student to appreciate the limited explanatory power of causality to capture the essence of local variation. At the same time, I ventured that this environment would allow students to use causality to articulate features of distribution. The results of the third iteration showed that, at the micro-level, causality could not properly explain the random effects. At the same time, at the global level, causality was harnessed to articulate the relationship between parameters in the model (average, spread) and the shape of the distribution. I regarded this paradox of seeing the limitations of causality at the micro level while recognising the power of causality at the macro level to be at the heart of coordinating the two perspectives on distribution. I asked whether and how students can coordinate the *data-centric* and *modelling* perspectives on distribution, and I elaborate this aim in this iteration. The results of the third iteration supported my conjecture that it is possible to design an environment in which students' well-established causal meanings can be exploited to coordinate *data-centric* and *modelling* aspects of distribution. Students (aged between 14 and 15 years) appreciated how, in addition to themselves as agents of variation, randomness instantiated in the form of the quasi-concrete arrows, can create histograms in which variation is apparent. In this sense, randomness might become understood as reality once removed. What I have called "letting go of determinism" might be seen as delegating control to a quasi-concrete object that exercises that power through random effects. The process of developing a model that can be used systematically to test out my conjecture made me particularly aware of introducing a graphical representation of the *modelling* distribution accessed by clicking on the relevant variable such as angle or speed (see Fig. 1). Once the button of the relevant variable was pressed a dialog box showed the distribution of values from which the computer would randomly choose, given a particular value for the slider as set by the student. The students moved either the arrows or the handle on the slider and observed the impact of their actions on the graphical representation of the *modelling* distribution. The microworld allowed students the facility to transform the *modelling* distribution directly but the *data-centric* distribution indirectly. While the simulation was

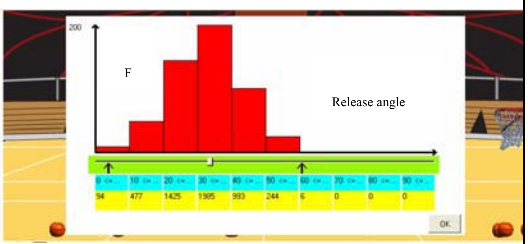


Fig.1: The microworld affords students the opportunity to change the way that the computer generates the data, by moving either the arrows or the handle on the slider.

playing, the students had access to both the *modelling* distribution and the *data-centric* distribution and they were asked to compare the graphical representations of the two perspectives on distribution. In this iteration, the microworld was used by eight pairs of students (aged 14–15 years) in a UK secondary school. Typically students of this age will have only encountered distribution as a collection of data generated from an experiment. I focus on the work carried out by two pairs of students as they engaged in making connections between the *data-centric* and *modelling* perspectives of distribution. I captured their on-screen activity on video-tape and transcribed those sections to generate plain accounts of the sessions. Screenshots were incorporated to make sense of the transcriptions. Subsequently, I analysed the transcriptions in attempts to produce extended narrative accounts for the students' actions and articulations. The excerpts in the next session are taken directly from transcriptions of the video tape. ('Res' refers to the author.)

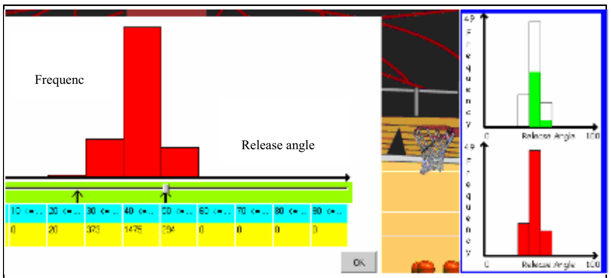


Fig.2: The modelling distribution (graph on the left) shows how the values of the corresponding variable were chosen. The graphs on the right plot the data generated by the computer, divided into successful throws (top) and unsuccessful throws (bottom).

FINDINGS

Below, I outline the interactions of Anna and James' with respect to the two perspectives on distribution. After Anna and James had recognized that the modelling perspective and the *data-centric* perspectives of distribution were showing different things, they were asked to compare the shapes of the graphical representations of the two distributions (Fig. 2).

Anna: Ehm … They are similar, but not the same, because they both have got the tallest bar and then two shorter …

Res: How can they be similar and tell different things?

James: Because, they've got the tallest in the middle, the second tallest on the left and then the second tallest on the right ….

Their reaction at this stage was to recognize that isolated features or isolated bars of the *data-centric* distribution were similar to their corresponding bars of the *modelling* distribution. This reaction was followed by considering the shape of a distribution as if it were an accumulation of just a few isolated bars: "the tallest in the middle, the second tallest on the left and then the second tallest on the right." A few minutes later I suggested that they look once more at the histograms.

Res: Are their shapes becoming the same?

James: Yeah, because that one is showing you what angle you selected (he was talking about the *modelling* distribution)

Res: Yes, show me with the mouse.

James: (He pointed to the red graph on the left). This is showing what you selected and then these angles … they are showing which one you are using.

Res: Yeah … but the shape as we can see is gradually becoming the same … Is it a coincidence?

James: No, It's meant to do that.

They began to articulate an understanding of the *modelling* distribution from which the histogram of frequency of successes against angle generalizes. This is articulated in terms of intentionality. But the sense of intentionality remains insufficiently clear. We can have at least two different possible interpretations: a) the intention is simply an expression of the pre-programmed deterministic nature of computers- at least in their experience; b) intentions are reflections of the actions of a modelling builder. Let us examine the data further.

Res: It's meant to do that. Why?

James: Because, the computer ….

Res: The computer?

James: the basketball man … because he has been told to use that angle the most (40-50) and then he is told to use this angle (30-40) and he is told to use this angle (50-60) … so, that's what it meant.

Res: Do you agree with him that the computer man is told to use those angles?

Anna: Yeah.

Res: Ok …Why wasn't he told to use those angles from the very beginning?

James: Because, he did not know that.

Res: And who has informed him about which angles to use?

James: We did …

Res: When did you tell him which angles to use?

James: By changing the slider.

I believe that the above protocol shows a transference of agency from the computer towards their own actions. This can be interpreted as a search for causes of variability and it might portray students' intrinsic need to shift to the problem of inferring causality by resorting to a causal explanation. James and Anna did not comment on the differences and similarities of the two perspectives on distribution and I wished to probe into their understanding of this aspect.

James: Similarities are that the tallest one, the tallest bar on the middle. The second tallest bar is on the left, and … the last one is on the right … ehm … the difference, is this one … it tells you (showing the red bar on the left) … this tells you what angle you selected, which one he gonna use the most and this one (the red graph on the right) tells you how he used all together in each throw, and this shows you the success rate of all the throws.

Res: Is their shape the same?

James: Yeah.

Res: So, the basketball player was meant to play as he did.

James: Because, you told him to. Well, it's meant to play between those lines, but … on that slider.

When the students talked about the similarity of the two graphs, they did so in terms of the relationship between the heights of the bars. In contrast, when they referred to the differences, they appealed not to the specific data, which was self-evidently different form bar to bar, but rather to the underlying role of the two distributions. They referred to *modelling* as what was intended but to the data distribution as what actually happened. James talked comfortably about variation which was the *data-centric* perspective. At this point, the vicissitudes of randomness, and the lack of a strong sense of the probabilistic mechanism, had prevented James and Anna from talking explicitly about chance and randomness. I think this is attributed to a lack of a clear probabilistic-type language for talking about randomness and probability.

I now wish to describe an alternative perspective articulated by Nick and Sarah. They expected an approximate equality between the *modelling* distribution and the *data-centric* distribution; after a certain number of shots, they expected the two distributions to be equal.

Sarah: Eh …that (showing to the graphs on the right hand side and the graph on the left hand side)

Nick: Not basically … Basically that (the graph on the left hand) he hasn't took a shot from there yet, so it's not calculated it (showing the red graph on the right hand side)

Res: So, do you think that these two red graphs are going to be the same?

Nick: Yeah … they will be … it's just this bit (pointing to the interval of 20 to 30, which it hasn't appeared yet on the red graph on the right hand side).

Res: Would they be in the future?

Sarah: Yeah, because it has not taken a shot from there yet … that's why it's not.

Res: What will make the two graphs the same?

Nick: When it will take a shot from 20 to 30.

They certainly expected isolated bars to appear on the *data-centric* distribution when certain shots would be chosen from the *modelling* distribution. A few minutes later, I suggested that they look once more at the graphs (Fig. 3).

Nick: They are similar…

Res: Are they exactly the same?

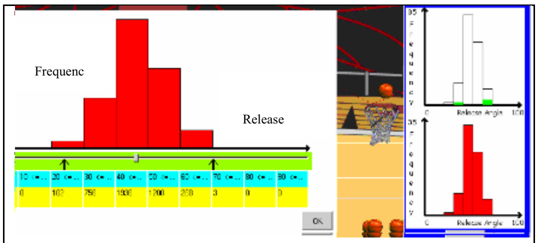


Fig.3: I asked them to compare the graph on the right (*data-centric* distribution) with the graph on the left (*modelling* distribution)

Nick: No … not exactly.

Res: Do you think they will be exactly the same at some point?

Nick: Yes.

Res: Why?

Nick: Because now it's got that bit (showing to the graph on the right) and they will become the same … this (pointing to the yellow part) is calculating what it's gonna look like after that many shots.

Res: Aha … aha … what is the difference between these two red graphs?

Nick: Because it's getting closer to the number … so, when … that … when the number of balls, he shoots goes to 102 that it will stay (pointing to the 20-30 interval of the graph on the left hand side) for the rest of it.

Res: Yeah.

Nick: And then he is taking 756 shots that (pointing to the interval of 30-40 in the *modelling* distribution) would be there (pointing to the *data-centric* distribution)… and when he is taking 1906 shots that would be there and then he is taking that amount of shots, would be that.

Res: So, what you are trying to say is that …

Nick: They will become the same.

Res: What you are trying to say is that the more times he throws …

Nick: The closer they will become.

The subjects never articulated a clear probabilistic appreciation of how the shots were generated. Nevertheless, they were able to express the notion of the *modelling* distribution as a target for the data distribution. It seems that perceiving the *modelling* distribution as a target towards which the *data-centric* distribution is directed is not dependent on understanding probability. Nick read the numbers of shots that would be chosen by referring to each interval in the *modelling* distribution and predicting that the corresponding number would appear in the *data-centric* distribution. Nick tried to explain why the two distributions would be the same.

Nick: They are the same, If you could that (showing to the 40-50 bar of the graph on the right hand side) on this (showing to the 40-50 bar of the graph on the left hand side) It will say … ehm … it will say … ehm … it will say like "number of results between 40 and 50 degrees is 1906".

Sarah: Yeah, eventually … yeah when …

Nick: and they will look the same … they will be the same ….

Sarah: If it will wait to 102 which is the smallest one, they might seem the same

Nick explicitly recognized that the frequency of the 40–50 class on the data distribution would be equal to its corresponding frequency on the *modelling* distribution. It is all the more striking that the factor of equality of the two perspectives is present from the prediction which Sarah made at the end of the fourth task. But Sarah was able to foresee the equality of the two graphs only when the 20-30 class interval for both distributions was 102. Her prediction concerning the bar which displayed that interval was quite curious since it elicited that prediction of equality. A few minutes later, the students articulated a revealing remark:

Nick: The more he tries, the more …

Sarah: The more time it takes, the more the angle … the more the graphs will look the same.

Students had some sense of bridging the two perspectives, and this progress is built upon a growing appreciation of how the Law of Large Numbers regularises the *data-centric* distribution, giving it shape and substance. However, their expectations do not show the underlying probabilistic mechanism by which shots are chosen from the *modelling* distribution to generate the *data-centric* distribution.

DISCUSSION

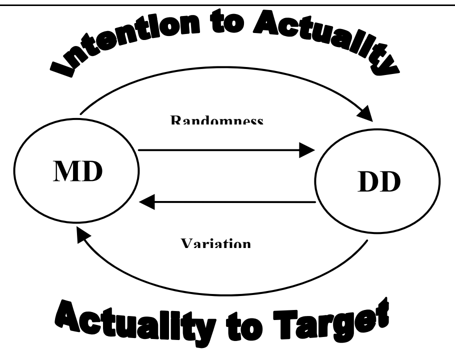


Fig.4: A tentative model for the connection of the data-centric and modelling perspectives on distribution.

After the subjects showed they were able to comprehend the role of the model and its different features, and distinguish the model (*modelling* distribution) from the real data (*data-centric* distribution), they were asked to compare the two perspectives of distribution. The students made an intuitive synthesis of the *modelling* and *data-centric* distributions, schematised by the structural model in Fig.4. The model shows that the students can perceive of the *modelling* distribution (MD) as the intended outcome and the data distribution (DD) as the actual outcome. Students can also perceive of *modelling* distribution (MD) as the target to which the data distribution (DD) is directed. Nevertheless, neither the intentionality model nor the target model is dependent upon a strong appreciation of probability. The target and intention models are not dissimilar perhaps from how experts appreciate the co-ordination of the two perspectives on distribution. However, experts add to this image the probabilistic mechanism underlying the relationship between the two perspectives.

REFERENCES

Ainley, J., Pratt, D. and Hansen, A. (2006). Connecting engagement and focus in pedagogic task design, *British Educational Research Journal*. *32*(1), 21-36.

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal, 3*(2), 64-83.

Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions, *Statistics Education Research Journal*, *3*(2), 42-63.

Biehler, R. (1989). Educational perspective on exploratory data analysis. In R. Morris (Ed.), *Studies in mathematics education* (Vol. 7, pp. 185-201). Paris: UNESCO.

Cobb, P. (1999). Individual and collective development: The case of statistical data analysis. *Mathematical Thinking and Learning*, *1*(1), 5–43.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research, Educational Researcher, *32*(1), 9–13.

Harel, I., & Papert, S. (1991). *Constructionism*. Norwood, New Jersey: Ablex.

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.

Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New Approaches to Numeracy* (pp. 95-137). Washington: National Academy Press.

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

Papert, S. (1996). An exploration in the space of mathematics educations. *International Journal of Computers For Mathematical Learning*, 1(1), 95-123.

Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children.* (L. Leake Jr., P. Burrell, & H. D. Fishbein, Trans.). New York: Norton. (Original work published in 1951).

Pratt, D, (1998). *The construction of meanings IN and FOR a stochastic domain of abstraction.*Unpublished Doctoral Thesis, Institute of Education, University of London.

Prodromou, T. (2004). Distribution as emergent phenomenon. *Proceedings of the British Society for Research into Learning Mathematics*, *24*(1), 49-54.

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 Sciences*, *5*(2), 69-88.