

WORKING GROUP 5 Stochastic thinking

CONTENTS

Working Group 5 on “Stochastic Thinking” <i>Dave Pratt</i>	481
Invited paper Probabilistic and statistical thinking <i>Manfred Borovcnik</i>	484 485
Papers on Statistics Education A structural study of future teachers’ attitudes towards statistics <i>Carmen Batanero, Assumpta Estrada, Carmen Díaz, José M. Fortuny</i>	507 508
Teachers’ representations of independent events: what might an attempt to make sense hide? <i>Sylvette Maury, Marie Nabbout</i>	518
The nature of the quantities in conditional probability problems. Its influence on problem solving behaviour <i>M. Pedro Huerta, M^a Ángeles Lonjedo</i>	528
Exploring introductory statistics: students’ understanding of variation in histograms <i>Carl Lee, Maria Meletiou-Mavrotheris</i>	539
Improving stochastic content knowledge of preservice primary teachers <i>Robert Peard</i>	549
Randomness in textbooks: the influence of deterministic thinking <i>Pilar Azcárate, José M^a Cardeñoso, Ana Serradó</i>	559
Papers on Probabilistic Thinking Problab goes to school: design, teaching, and learning of probability with multi-agent interactive computer models <i>Dor Abrahamson, Uri Wilensky</i>	569 570
Strengths and weaknesses in students’ project work in exploratory data analysis <i>Rolf Biehler</i>	580
Randomness and lego robots <i>Michele Cerulli , Augusto Chiochiarriello , Enrica Lemut</i>	591
Students' meaning-making processes of random phenomena in an ICT-environment <i>Kjærland Iversen, Per Nilsson</i>	601
Young children’s expressions for the law of large numbers <i>Efi Papanastasiou</i>	611
Towards the design of tools for the organization of the stochastic <i>Dave Pratt, Theodosia Prodromou</i>	619

WORKING GROUP 5 ON “STOCHASTIC THINKING”

Dave Pratt, *University of Warwick, United Kingdom*

This section of the proceedings reports on the work of the Working Group 5 on Stochastic Thinking, which incorporated issues pertaining to the teaching and learning of both probability and statistics, as well as the interface between the two. This group was led by four organisers: Dave Pratt (UK, Chair), Rolf Biehler (Germany), Maria Meletiou-Mavrotheris (Cyprus) and Maria Gabriella Ottaviani (Italy).

After an initial review in which all received papers were inspected by at least two other contributors to the working group, 16 papers were accepted for presentation within this working group, including one invited paper, presented as a plenary discussion. This initial review placed heavy emphasis on inclusivity and where possible reviewers gave sufficient support to enable the papers to reach a standard appropriate for initial presentation. The papers represented work from authors spread across three continents and eleven countries.



Fig 1: Michele Cerulli (centre background) leads an ice breaker exploring randomness through robots.

The working group began with an ice-breaker (see Figure 1), planned and led by Michele Cerulli (Italy), who set up several experiments, each involving the use of robots. Small groups of delegates worked on each task in turn, finding that in each case interesting questions were raised about the relationship between randomness and determinism. More detail about these ideas is provided later in the paper by Cerulli et al. The ice breaker provided opportunity for delegates to meet and talk, a crucial foundation we felt for subsequent

productive discussion.

There was one special session in which we invited Manfred Borovcnik (Austria) to talk more generally about the nature of statistical thinking. The other two themes, Probability and Statistics, each occupying two sessions. In each session, papers relevant to that theme were briefly re-introduced with the aim of reminding the group about their key ideas. The aim was that most of the session was devoted to asking specific clarification questions or raising important discussion points. We felt that that aim was successfully accomplished. Difficulties with language were handled through the informal use of interpreters where necessary.

Working Group 5

In order to summarise the discussion, we collected a set of research questions, which we would like to see the stochastics community address in its future work. The following questions were presented at the final plenary by Rolf Biehler.

1. Which factors influence negative and positive attitudes and values and sustainable interest towards probability and statistics (in students, pre-service and in-service teachers)?
2. How can mathematics teachers learn the specificities of probabilistic and statistical thinking as compared to mathematical thinking?
3. How can we support teacher development in using innovative pedagogies to teach stochastics in a new form that is more related to mathematical and statistical literacy?
4. How can we design better classroom studies to investigate the relationship between pedagogy (classroom culture) and learning outcomes?
5. What are students' situated understandings of basic concepts (such as: average, spread, distribution, determinism, causality, randomness, stochastic & physical independence) and how can we support their development?
6. How can we design new tasks, textbooks, computational environments, experimental environments to offer intuition-based pathways towards more sophisticated understanding?
7. How can students' experiences in various probability-and-statistics simulations be fostered towards more generalisable knowledge?
8. How does the setting (material, physical or virtual experiments, simulations) shape stochastic thinking? What are the cognitive affordances of virtual (on-screen) and material stochastic objects, and what are the design-and-learning issues related to this?
9. Respecting multi-disciplinary approaches to Exploratory Data Analysis, how should we investigate the potential tradeoffs inherent in introducing and incorporating realistic data and iterated data-analysis cycles?
10. How can assessment approaches for statistical competence reflect the complexities of students' activities (e.g. the role of project reports and answers to interpretation and data analysis tasks)?

After the conference, a second review phase was set up. The reviewers for this second phase were the four organisers of the working group together with Manfred Borovcnik, Robert Peard (Australia), Arthur Bakker (The Netherlands), Joachim Engel (Germany) and Katie Makar (Australia), and we would like to thank all these people for their efforts but especially the latter three, whose efforts were entirely altruistic, given that they were not involved in the working group at an earlier stage.

Working Group 5

The second phase of the review represented an opportunity for further improvement in the papers and we believe that, as a result of the dedication of the authors and the reviewers, the 13 papers all satisfied the increased emphasis on quality reflected in this second stage.

At the end of the conference and the subsequent collation of the proceedings, there was a feeling that the community established at CERME 3 has continued to flourish and will be able to support research in stochastics in the future. Indeed, we look forward to the continued work of this group at the next CERME conference.

Working Group 5

Invited Paper

Manfred Borovcnik

PROBABILISTIC AND STATISTICAL THINKING

Manfred Borovcnik, *University of Klagenfurt, Austria*

Abstract: *Mathematical concepts enable us to structure our thinking; corresponding models help us to structure reality. They supply us with tools to recognize and solve problems. Stochastic models are not mere images of reality that fit more or less. Right from the basics they have more the character of scenarios to explore reality. Given this circumstance, and the way that feedback about success is only indirect, understanding of concepts and the reasonable application of models is impeded. Accordingly, misconceptions are abundant and recipe application is ubiquitous. Stochastic thinking seems to be quite different from other types of thinking like causal thinking, or logical thinking. The educational discussion until the 90's coined the notion of 'probabilistic thinking', from the 80's the discussion shifted to the notion of 'numeracy' and 'statistical thinking'. By examples and figurative deliberations a multi-faceted image of probabilistic and statistical thinking will be given.*

1 Thinking in scenarios – some examples

The scenario feature of probabilistic thinking will be illustrated by some examples, which will also shed light on the merits of the probabilistic approach.

Transparency of decisions

In the face of uncertainty, a single decision may be made more transparent if one allows for weighing the various possibilities. The competing decisions can then be compared by computing (expected) values instead of actual costs or wins.

The problem dealt with here is whether one should take out a policy for a comprehensive insurance of one's car for the next year. The focus is not on mapping the situation precisely onto a model but on illustrating matters; the rough model should just highlight the situation and the purpose of modelling by probabilities. The following table will give the costs of the various decisions (insurance yes or no) under the prospective circumstances (no accident at all, total wreckage).

Cost [in Euro]		Decision	
		A ₁ = Insurance yes	A ₂ = no
Potential future	T ₁ = No accident	1000	0
	T ₂ = Total wreckage	1000	20 000

With hindsight, one can easily tell if it were better to take out a policy – if no accident happened, no insurance = A₂ is the better decision. If one minimizes the maximal cost, then A₁ = insurance is better. This corresponds to risk-avoiding behaviour. More

margin for innovative behaviour will be opened by introducing probabilities: if one is ready to 'weigh' the possible futures T_1 and T_2 , e.g. by relative weights of 39 : 1 (this corresponds to a probability of 1/40 for the total wreckage), then the cost of decision A_1 still is 1000, but the cost of A_2 has decreased to 500. Hence it would be better not to take out a policy. Clearly, the actual decision will depend on the weights for an accident. Other weights will lead to other decisions. However, the decision is now transparent: if one can weigh one's chances of such an accident by 39 to 1, then A_2 would be better. To free oneself of the burden to fix one's chances, it could be advisable to find the so-called break-even point, i.e. those relative weights at which the decision turns from A_1 to A_2 . Here it is 19 to 1. If someone evaluates his/her own chance to be higher or lower than this break-even, he/she should decide accordingly. With the latter procedure there is no need to weigh someone's chances exactly.

Judgment of risks

Not only technical systems have a reliability of survival dependent on its components. One may derive probabilities for the whole system to operate from an elementary (or more sophisticated) assumption. The result has more or less the character of scenario figures and gains more relevance in the comparison of various changes to the system. This will give indications for which changes to promote.

The following problem is drawn from engineering applications. A system has 3 components, the reliability of each is 0.95 for a specific purpose – e.g. that they are working well for exploring the Titan in a special mission. The system works if B_1 and B_2 operate well, or if B_3 works, see Fig. 1

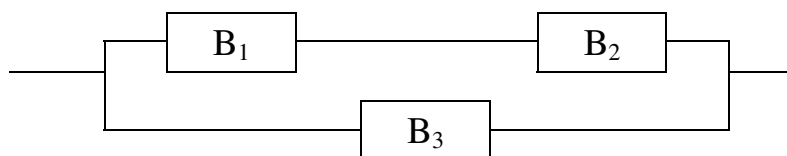


Fig. 1

What is the reliability, i.e. the probability of working well for the mission of the whole system? Is it better to take two full systems on the mission, or is it better to have each component doubled? How many complete systems or, how many stand-by components for each one should be taken on the mission, if the reliability of the resulting system is required to be 0.99999 (whatever that should mean)?

The standard solution treats the components as if they were independent of each other and have really the same reliability. Of course, they are *not* independent, and they have *not* the same reliability, and their reliability is *not* 0.95 (this is a *qualitative* statement of the involved engineers). Yet, the scenario is the only way to deal with the problem before the spacecraft is sent on its mission to the Titan. From this, one gets an idea of the reliability of the final system and whether it would pay the additional costs to build in a specific level of redundancy. A residual risk will always remain; the scenarios, however, will allow judging relative risks and costs of the various actions.

Fixing prices in the face of uncertainty

Expected values are the basis of fixing prices when the future is open to variation. The procedure necessitates weighing the various possibilities. One basis for weighing is taken here, namely to extrapolate risks of the past to the future, i.e. to use relative frequencies for the estimation of probabilities of the various risks at hand.

For a single car-owner e.g., the relative weights of total wreckage have to be measured individually; the insurance company may rely on the statistics of accidents of the last years. For the sake of simplicity, we will again assume only two possibilities for all the policies: total wreckage and no accident. With 2% total wreckages from the past and 10 000 policies, the bookkeeping of the insurance company looks like:

200 total wreckages with a cost of 20 000 each amounts to payments of 4 000 000, i.e. 400 Euro per policy. This amount plus an equivalent for taking the risk plus expenses plus profit makes ..., let us say 1 000 Euro per policy.

However, there is always a remaining risk for the company (not in our modelling here but in general) that the premium will not be sufficient. How high will be that remaining risk? How will it change if there are only 100 policies, or if there are 100 000? The higher remaining risk with a smaller number of policies reflects the fact that the (implicit) use of expected values is less reliable in small companies/samples. How will the risk change with various levels of the premium? Are all possibilities included in the scenario? No, the scenario is suitable only for a normal financial year, for the occurrence of catastrophes, the insurance companies have the system of reinsurance, which reduces the remaining risk (of making high losses) by distributing it to a higher number of policies.

Concluding from circumstantial evidence

At court, if no confession is available, the judges have to rely on circumstantial evidence. Doctors diagnosing various diseases have to rely on indications from blood tests, X rays, mammography results etc in order to decide about medication or operation of a patient. There are ubiquitous situations where conditional probabilities could help to find the direction of further measurements to be taken best. Formal calculation is done according to Bayes' formula.

In what follows, we will refer to a blood test for diagnosing HIV with the following reliabilities: a person with the virus will be recognized by the diagnosing procedure (test positive) with a probability of 0.99 (in medical jargon this is called the sensitivity); a person not having the virus will be judged virus-free (test negative) with probability of 0.987 (the rate of false positives therefore is 0.013). For a person with positive result, how high is his/her risk to actually have the virus? If we apply the scenario to a person being representative of the whole population, we could use the prevalence of HIV (e.g. 0.02%) and come up with a probability of 0.0150 for having the virus under the condition that the test was positive. Judging the person to belong to a high-risk group with a prevalence of 10%, will result in a probability of

0.8943. For a discussion of the impact of various input probabilities on the interpretation of the resulting probabilities, for example on the question of how to communicate these to the patients, see also Gigerenzer et al. (1998).

How do we perform an adequate calculation of the actual probability of having the virus after a confirmation test when the first result was positive and the second then also was positive? Can one simply combine two applications of the Bayes' formula? (There is evidence that the reliabilities of the test conditional to the first positive result are not the same as before – in other words, if the test has wrongly gone positive it will more likely do the same a second time).

Which scenario is more applicable to the patient? Where do the reliabilities of the testing procedure come from? Do they also have only figurative character or have they come from a controlled experiment with blood samples of which the status of HIV (or not) was absolutely clear? Do we recommend the testing procedure for mass screening? What will be the consequences of mass screening? Is the testing suitable as diagnosing procedure? How can one improve the diagnosing procedure?

Again, the application has somehow the character of a scenario and gives more information about which action to do next. At the level of implementation of the diagnosing procedure e.g. it will allow a transparent deliberation of advantages (more precise information) and the relative 'cost' (from wrong positives and from wrong negatives). For a teaching approach including such questions, see Vancsó (2003, 2004). For an example related to mammography and resulting doubts if it should be introduced as screening procedure for 50+ women, see Hoffrage et al. (2000).

2 Approaches to generalizing information

Singular data sets are prone to variation and therefore could convey everything. If someone seeks to adapt to future events, or if someone wants stable descriptions of the 'status', then what to rely on? There is a big need, there is also a desire to extract general features from singular data sets, i.e. to *generalize* the results found. Accordingly, there are a lot of strategies and models for this purpose. If one can derive other statements from the data at hand by logical argument, fine. If one can find out the exact conditions that will lead to a specific (desired) result by causal connection, fine. If though the results are also due to some yet unexplained but *small* variation how does one find out that?

L'homme moyen

This figurative idea of Adolphe Quetelet (1835, see Stigler 1986) is intended to explain how a person gets his/her final outcome of a characteristic (e.g. height or head circumference etc) by a value that represents the *l'homme moyen*. By errors of nature, however, this value is superimposed by many small errors.

The *individual* process of superimposing small errors to the (ideal value of) *l'homme moyen* 'leads directly' to the normal distribution of the investigated characteristic *in the population*. Quetelet transfers herewith the mathematical model in the

background from the elementary error hypothesis from physics to a very broad range of applications; the derivation of the normal distribution is based on the central limit theorem of Laplace, which was well-known at that time. Historically this marks an early milestone of the application and interpretation of the normal distribution outside the error theory of physics. The enthusiasm about this transfer was so great that at times the normal distribution was rated to be a *law* of nature.

$$\boxed{\text{data}} = \boxed{\text{l'homme moyen}} + \boxed{\text{many small errors}}$$

Remarkably, the addition of these errors amounts to the unexplained variation of modern views, which is *modelled* by randomness (a normal or some other distribution). The l'homme moyen represents the model for the data. In terms of generalization, it is the generalizable feature of the data, which is estimated by filtering out random fluctuations from the sampling process.

Other structural equations to split the data describe the generalizable part differently. All have their own interest:

$$\boxed{\text{data}} = \boxed{\begin{array}{c} \text{signal} \\ \text{pattern} \\ \text{fit} \\ \text{model} \\ \text{explained} \\ \text{common} \end{array}} + \boxed{\begin{array}{c} \text{noise} \\ \text{deviation} \\ \text{residual} \\ \text{residual} \\ \text{unexplained variation} \\ \text{specific causes} \end{array}}$$

All these equations describe different approaches to model the individual's deviation from a true or theoretical value (representing the generalizable part of the data), which may be re-interpreted commonly as l'homme moyen to facilitate understanding. Some of these approaches model the deviations of single objects by random fluctuations (see the classical inference approach below), some by causal influences from other sources, some use a mixture of both (see the ANOVA approach below), while others analyze the deviations by mere patterns and explanations of the patterns from the context of the data (see the EDA approach below).

Classical inference from data

Observed data are usually summarized to give

- predictions for future events
- a generally valid description of the population – e.g. a confidence interval

Both procedures include a risk of statements not being valid and necessitate estimating the magnitude of the variation. Furthermore, they require that the data generation process is a *random* sample of the population. In terms of the idea of

l'homme moyen this means: If one knows the value of l'homme moyen and the magnitude of errors, i.e. the variation, then this amounts to generalizable information, i.e. one can predict future outcomes to be within the following bounds:

$$l'homme\ moyen \pm variation$$

If a person is within these bounds then it is due to no special cause, only *natural* variation due to random sampling is effective; if not then other specific (usually causal) explanations have to be searched for as he/she does not fit to the general case. In other words, there 'must be' some other l'homme moyen type working for that person.

This pattern of argument follows the modern interpretation of the concept of distributions. The normal distribution (as other distributions) is interpreted as an external phenomenon, i.e. as a tendency to produce (by random sampling) an individual with a specific measurement of the investigated characteristic. The underlying internal process of causal or other factors influencing the final value is usually no more an object of study (with the exception of the ANOVA approach in empirical investigation, see below). Usually these other influences are balanced and eliminated for a whole group of observations. However, there are some cases for which this balancing is not effective, which might make a causal interpretation of these influencing factors relevant.

Actually, Wild and Pfannkuch (1999) find out that the tendency for searching for such specific causes is very deep-seated and would lead people to seek for causes (for shifts) also in case that an individual's data are quite within the predicted bounds based on a (pure) random model for the variation. For example in sports, if the 'scores' decrease, the trainer or other responsible persons of a team are inclined to look for the special causes for that development even if a formal analysis of current scores does not yet yield a *significant* decrease of achievement. This gives a more direct basis for earlier intervention e.g. the initial decline in achievement say in sporting prowess (but also in a quality control setting). That means that people are more ready to search for a causal re-interpretation of observed data than to model them solely by pure randomness. And in a way they are sometimes quite right in their approach.

If the process of data generation does not compare to the *naturalistic* process of elementary errors acting upon the l'homme moyen, then the data cannot be generalized and used in the way indicated before. In modern terms, we could say that in this case the data are not from a random sample of the population. The key to allow for generalizing findings from data is that they stem from a random sample.

Analogous deliberations may be made for the confidence interval method to generalize findings from data. Clearly, the context will be important to judge whether the random sample argument is valid and if the sample is actually taken from the target population to which the findings are to be generalized.

EDA approach towards inference from data

Exploratory data analysis is centred on the following structural equation:

$$\boxed{\text{data}} = \boxed{\text{pattern}} + \boxed{\text{deviations}}$$

So-called robust techniques should allow filtering out the pattern which means the procedure to find patterns should not be affected too much by some unusual, extreme data. Here, we are farthest from the idea of Quetelet. There is no process of natural superimposing of small errors; they could even be very different to different elements of the population and quite big sometimes. In terms of the classical approach, there is no need for the sample to be random.

The justification for generalizing the split of data into a specific pattern and deviations comes from the knowledge of the context: the pattern should give an interesting insight into the context; the deviations may sometimes even shed more light onto the problems within the context. Furthermore, the aim of EDA is a multiple analysis of the data with the aim of several splits and resulting views on the underlying phenomena. The power for generalizing the findings is based on the *comparison* of the various patterns found. Deviations get more attention; they do not merely reflect small error entities ‘caused’ only by nature. Accordingly, a lot of effort is also invested to interpret those deviations on the basis of context knowledge and explain or see why they are as they are. In terms of the ‘cause split’ of data, EDA looks for common causes in the pattern *and* for specific causes in the deviations.

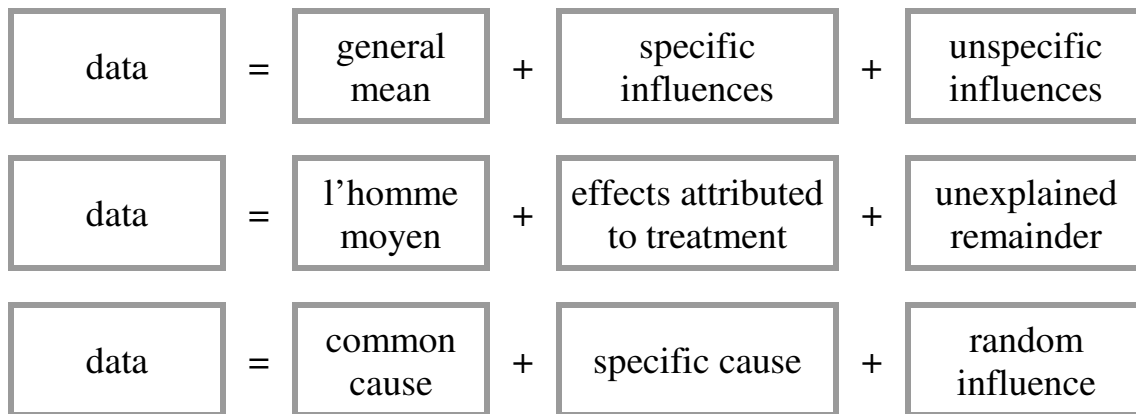
Of course, different sides in a ‘conflict’ would evaluate context knowledge differently. There is a lot to undertake to improve the preliminary and/or subjective character of results generalized from one data set. Sometimes further projects produce more data giving more conclusive results; or a cross-validation of potential results with subgroups, or with other co-variables, or with other populations or similar problems could help. Also the EDA approach may be shifted to the inferential approach in the ongoing phases of analysis to corroborate preliminary findings.

The ANOVA approach to generalize findings from data

The analysis of variance approach is a standard but sophisticated technique to split the variation in the data into that from specific sources and the rest variation that is modelled by randomness (usually by the normal distribution). We will not go into technical details here (see e. g. Montgomery 1991) but will just discuss the structural equation for the data and its similarities to the l’homme moyen idea.

For illustration, we will use the context of various teaching methods A, B, .. (= treatment) which might have an influence on some achievement score – the target variable of which data are available. The specific influences are attributed to the treatment A (or B, ..) a ‘person’ has actually got, the unspecific influences are modelled by randomness. The formal procedure of ANOVA allows deciding when the variation due to the specific influences is big enough as compared to the variation

due to randomness in the remainder, so that an influence of the treatments may be generalized from the data.



With respect to Quetelet's figurative thinking, we separate the deviations of l'homme moyen into two parts, one is causally explained by the effect of the actual treatment, and a remainder that is not yet open to a causal explanation and that should also be small enough that it would not pay to search for further causal explanation.

3 Structuring of thinking

Concepts and models allow us to 'see' reality in a specific way – this acts as feedback also on our thinking; it structures our thinking insofar as it anchors analogies and figurative ideas and archetype models in our approach to reality. We have seen from the examples in section 1 that probability has a strong feature of building up scenarios for reality which means that usually it does not directly model reality in the sense of constructing a model as a more or less good image of the real situation. As a consequence, that has a deep impact on how learners can integrate probabilistic concepts into their repertoire as there is only indirect measures of how close the model depicts the real situation and also the *success* of a model is not easily judged. No wonder that misconceptions are abundant and deep-seated, i.e. are often not revised by pertinent education.

Feedback in probability situations is only indirect

Normally, we learn by trial and error. By wrong decisions we lose and thereby we think about improvements and come up with better models and so forth. Here only one simple example is given to illustrate that matters with probability are much more complicated and feedback about success is by no means open to direct interpretation.

Take the two *Falk wheels* of fortune of Fig. 2 and give the choice which to spin. Obviously the left wheel has the bigger sector for winning (=1), therefore one should choose it. However, once you spin, you have not a high probability of winning, .. , and you might lose. What then? Was the choice wrong? The best action here may not be awarded. Thus, you might speculate about the reasons why. At times, I checked this with my then seven years old daughter and right from the beginning she wondered about the special margin of the right wheel, ..., and concluded that grown-

ups should take the left but children had to take the right...(the author named the wheels after Ruma Falk).

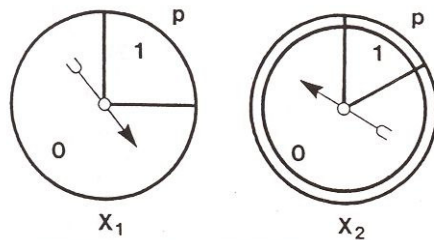


Fig. 2

With reference to the scenario use of probability in section 1 and with respect to the example above, relative frequencies as companion of probabilities are thus often more or less a metaphoric way to describe probability (for the genuine theoretical nature of probability see also Steinbring 1991). Yet they amount to a useful way to describe some features of the abstract concept, which is out of the reach of understanding. All the more relative frequencies often may be used as a short cut to a mathematical derivation of probabilities by the method of simulation.

The idea of weighing the evidence

Relevant information about a situation under uncertainty may come from:

- Combinatorial multiplicities of possibilities, which are judged to be equally likely
- Frequencies of events in past or comparable series, which are judged to be 'similar'
- Personal judgement of involved risks

Core of the further information process to derive at a decision is the concept of expectation as illustrated in some of the examples of section 1. Quite often the three types of information above involve – despite exact numerical values for the probability requested – *qualitative* aspects and the scenario character of the models used. Calculating (expected) values of tentative decisions in order to find a justification for the final decision made may be viewed as an exchange between 'money' (utility, cost) and uncertainty like in the car insurance example of section 1: The small probability/risk of an accident with a high loss of money is exchanged for the premium of the insurance, which is a small but fixed amount of money to be paid anyway in advance. The question is how to find a justifiable amount of money for the actual premium, or how to decide if one should take out a policy if it is offered for a specific premium.

If someone is faced with a decision, the consequences of which lie in the future and cannot be foreseen (except a listing of all cases, which are considered to be 'possible'), then one has to find some optimality principle to signify one or some of the decisions as better than others. Minimizing maximum 'loss' works without introducing probabilities; ordering decisions by expected 'loss' requires the weighing of uncertainty. The potential of the method increases by an investigation of how

sensitive the derived decision reacts to changes in the weighing process – again a plea for the scenario character of the probability approach.

Mis-Conceptions

With the indirect feedback to the probability approach, it is not surprising that thinking in probabilities is not well integrated into the cognitive frame of individuals. Often, modelling a situation involves an attribution and separation of and between causal and random parts (see the discussion in section 2). The causal part of a problem, if separated and explained, allows for more direct interventions and thus seems more promising.

For example, in Wild and Pfannkuch (1999), the actual score of 2 out of 5 is compared to a probability of scoring of 70%. ‘Has something gone to the worse, or is the team as good as always and had only an unlucky series, or can one find a specific explanation for the low achievement at present?’ According to the classical inference approach, modelling comprises only random elements and an appropriate statistical test yields that the achievement is not significantly low, which means it is within the usual fluctuation and therefore no intervention is necessary.

People, however, tend to seek for other, mainly causative explanations of the low actual score. Once one has found such a causative explanation (e.g. temporary private problems of the player, a small physical injury, a quarrel in the team etc), the track for success promising intervention is open. This in mind, people seem to extremely favour the causative approach as compared to the random approach (see Wild and Pfannkuch 1999).

In the tradition of Kahneman and Tversky (see e. g. Tversky and Kahneman 1972, or Kahneman e. a. 1982), a lot of misconceptions have been identified and described. Intuitive strategies like representativity, anchoring, and causal strategies constitute subsidiary strategies to surmount the difficulties in the process of weighing the uncertainty (for a discussion see Borovcnik and Peard, 1996). Among many others, here only the outcome orientation of Konold may additionally be referred to.

According to it, information available to a person is more likely to be actually used in solving a problem if it allows for a direct prediction of the outcome in quest, or, the problem is reformulated in order to allow for such a direct prediction (see Falk and Konold 1992): For example, the probability of 0.95 for rain allows for a direct prediction to encounter rain if one goes out – with a small risk of having no rain, which is even rated to be smaller than it actually is if it is not neglected at all. And if it actually does not rain, people would complain that the weather forecast was *wrong*. If the probability were 0.5 – fifty fifty – then people tend to pick up any external or causal information if it serves to predict the ‘outcome’ in question directly. This fits quite well to Wild and Pfannkuch’s observation above.

From the abundant misconceptions, one may conclude the difficulty of the venture to teach *probability* concepts and its necessity. This is also true for the education of the *statistics* part as not only Wild and Pfannkuch (1999) describe individual’s problems

to discriminate properly between causative and random parts in splitting and explaining variation in empirical data.

To use misconceptions effectively in teaching, it is not sufficient to confront learners with relevant situations which are prone to wrong approaches; instead of trying to revise wrong intuitions (which are very basic and deep-seated) one should build a bypass by re-presenting the situations by (very) simple material forms that allow for solutions. For promising examples in view of the Bayesian formula, see the unit square of Bea and Scholz (1995) representing (conditional) probabilities by areas, or the natural frequencies, see Krauss et al. (2002), or Hoffrage et al. (2002), representing (conditional) probabilities by absolute (natural) frequencies. For some other crucial concepts such basic material representations are still waiting to be 'invented'.

Probabilistic and statistical thinking

From the examples in the first two sections, types of situations and types of thought that help with them are derived. In all cases to follow, a mingling between probabilistic and statistical thinking may be traced. With some twist of thought, the one or the other part predominates. Always the following strategies may be supportive in presenting (or in solving) the following standard situations: To give simple analogies, which are similar in characteristic features and which have illustrative potential for the solution. To give simulations which necessitate organizing clearly the model assumptions for the involved situation and effortlessly yielding the solution. To re-formulate, or even to re-present the situation in a more basic manner.

One-off decision: The procedure of attributing (expected) values to decisions by an exchange between uncertainty and 'money' goes back to Christian Huygens 1657 (see Bentz 1983 or Freudenthal 1980). He developed his ideas in the context of lottery games and speaks of some uncertain lottery to 'be equally worth as' some amount of money given by a formula (for the expectation). Huygens himself already applied his concept of expected value also to insurances especially to life insurances. The less agreement there is on weighing the uncertainty, the more it gets important to supply additional justification to the weighing by investigating the consequences of different weights – the stronger becomes the scenario character of probability as was discussed in section 1.

To think about uncertain situations in terms of scenario like expected values and respect that as one tool amongst others to derive at transparent decisions is a basic ingredient of stochastic reasoning. It cannot be stated clearly enough that the *values* for the decisions allow to signify some decisions as better than others *without* making it possible to predict (or even attempting to aim at predicting) the specific outcome in the 'future' – this seems counter-intuitive not only against the background of Konold's findings on the ubiquity of the outcome orientation in individuals.

Decision in the face of circumstantial evidence: Abundant situations require proceeding according to the actual values of conditional probabilities relative to new facts. Judgement before court by circumstantial evidence, or diagnosing procedures in medicine, are just two prominent examples of that kind of reasoning. Formally, the new conditional probabilities are calculated with the formula of Bayes. The never-ending disputes about the validity of prior probabilities (prior to the circumstantial evidence) indicate again the scenario-character of probability. Furthermore, the two involved ‘directions’ of conditional probabilities do have a completely different connotation from a causal standpoint: Whereas the conditional probability of, for example, having some virus to having a positive result on the diagnosing test *is causally* (and it is only by some errors possible that a negative result is achieved at), the backward direction of conditional probability from a positive diagnosis to actually having the virus is *merely indicative*. This is a fact that is hard to accept for many due to a causative misinterpretation of it. Findings from research on misconceptions tell us that conditional probabilities are grossly overestimated in case of a possible causal interpretation, and are often even neglected (too small to be taken into consideration) when they are only indicative.

To think about pertinent situations in suitable terms of a Bayesian model, amounts to basic ingredients of stochastic thinking. There are many endeavours to improve teaching on these issues. While it seems inadequate to transform us to Bayesian thinkers, it establishes a great progress in the teaching of probability to make us aware about the cognitive biases from causal re-interpretations. A great help with that comes from very basic material re-presentations, which demystify the causal connotations that always come back to mind if the mathematics involved becomes too complicated. From all the endeavours, here only the approach of Krauss et al. (2002) of natural frequencies is referred to.

‘Natural’ variation of randomness: Investigating the variation of data involves often a split of the data into explained and unexplained parts, or in causative and random parts, see section 1. As this separation is neither unique nor clear-cut, our intuition has to be backed up by simulation studies about what it would mean if the variation were only random. What are the implications of ‘pure’ randomness? The so-called ‘square root of n’ law may be demonstrated by such investigations and teaching experiments, see e.g. Riemer (1991), or Kissane (1981): If a target variable is the sum – or better the mean value – of other variables (that need not necessarily have the same distribution, e.g. 1 or 0 according to some dichotomous experiment, or the result of throwing a die), then

- a two sigma-rule becomes more accurate with increasing number of summands: approximately 95% of (simulated) values of the target variable are between the mean value plus or minus two times the standard deviation of the target variable,
- the standard deviation of the target variable decreases by a factor $1/\sqrt{n}$, n being the number of summing variables.

This is a manifestation of the elementary error hypothesis of physics of earlier times and a concrete example of the central limit theorem that states that the ‘limiting’ distribution for the (mean) target variable is a normal distribution.

To think about pure randomness’ consequences in terms of a normal distribution and in terms of an ever-decreasing variation (by the square root of the number of ‘trials’) is the key to many a statistical procedure: to name only one, the estimation of a population mean by the mean of a random sample. The precision of that procedure is improved by larger random samples. The amount of improvement may be read off and adjusted to some accuracy required by investigating a sample, which is large enough. The model situation with the same summands (as in the simulation situation) being independently taken from the same population will also shed light on the importance of a *random* sample and not just an arbitrary sample to be taken in order that the law comes true.

4 Structuring reality

It would be too restricted to think of structuring reality by models as to depict the relevant features of a real-life situation, abandon less relevant ones, and to ‘filter out’ a model that represents more or less an *image* of the original situation. Within that model one could then derive mathematical solutions and re-interpret them into the context problem from the onset. This is *one* feature of probabilistic modelling only, one that is truly important, but there is more to say about probability models with respect to their more indirect and scenario character already described within the examples of sections 1 and 2.

Models and concepts allow us to structure *thinking* and this in turn allows us to apply these models to reality, structuring it and (partially) solving the problem there. It is worth devoting some extra thought on the objective side of models structuring *reality*. There are more ingredients that come to the fore with this focus, especially the interplay between causative and random parts of the variation of variables, which allows dealing with real problems.

An example for the structural equation ‘in action’

The following example deals with the explanation of the target variable ‘body weight’. From many examples we have learned that there is quite a tight relation between weight and height of persons. Another explanatory variable is gender. The remainder is unexplained (by further influential variables) and modelled to be random. It could, for example, be explained further by the body type of a person (pyknic, leptosome etc), or it could be explained by race, or by nutrition in early childhood etc.

$$\boxed{\text{Body weight}} = \boxed{\text{a constant value}} + \boxed{\text{gender effect}} + \boxed{\text{b} \times \text{body height}} + \boxed{\text{random 'error'}}$$

Without the gender term, one would have a simple regression line for the model to describe the relation between weight and height. Separated between genders there is quite a different slope of the regression line. If the investigator does not include gender into the study, the variable would have the status of a *confounding* variable. A confounding variable changes or even completely reverses the relations found. Once data are available on gender, the variable may be tested if it should be taken into the model; in case if it were a continuous variable like the height, it would be called *covariate*. A covariate is simply a *candidate* to be included in the model for the data.

The remainder is not open to causative explanation as there are no specific data available. It will be modelled simply by pure randomness. The evaluation of whether the explanatory variables should be integrated into the model for the data is done in comparison to the ‘size’ of the remainder: are the changes in variation, due to candidates for explanatory variables, big enough in comparison to the variation due to that remainder.

Neither is the interactive building of a model unique, nor are the components, which constitute the remainder, unique. For an interpretation of the final model for the target variable, the separation of variation into causative (explained) and random (non explained) parts is essential – the split is not ontological but only pragmatic. If this separation yields a relevant model, the causative interpretation may lead to promising interventions.

The formal procedure to separate the model entities is based on significance tests in the ANOVA or ANCOVA models; we will not go into details. An intuitive understanding of this separation is at the core of anyone’s reasoning who is involved or concerned with statements that are backed up empirically by data from investigations. Accordingly, textbooks on empirical research covering the ANOVA approach devote a lot of effort to develop it.

Furthermore, the intuitive strive for looking for causative explanations for phenomena at hand, is not being teased out by probabilistic reasoning. On the contrary, probabilistic models are a key factor to filter out causative elements of a problem to get more control over interventions on the target variable.

The split of variation into causative and random parts

In empirical research, often causal influences for the variation of a target variable are searched for. For illustrative purpose, the reader should think of several alternative treatments, which could affect a target variable. The simplest model would be (with some known function f):

$$\boxed{\text{Target variable}} = \boxed{f(\text{treatment effect})}$$

The variation of the target variable would then uniquely be determined by the treatment effect. However, there is a lot more sources of variation in the data for the

target variable like plain measurement errors, variable external circumstances of the experiment, other attributes of the ‘persons’ which also could influence the final value, variation due to the specific persons that are sampled and investigated, etc. We have here a similar but more differentiated situation as in the l’homme moyen figure of Quetelet. We will give a simple structural equation of how the target variable emerges (more precisely, the equation establishes only one simple model for the situation; there are many other suitable models):

$$\boxed{\text{Target variable}} = \boxed{\text{a constant value}} + \boxed{\text{treatment effect}} + \boxed{\text{influential variable}} + \boxed{\text{random 'error'}}$$

Treatment and *influential* variables are *explanatory* variables, which explain the variation of the target variable by some causative (or associative) argument, the random part could represent further causal relations between values of some other variables of the person investigated and his/her value of the target variable. If one can establish tight relations of some of these with the target variable, then they could be integrated into the explanatory variables (see also Wild and Pfannkuch 1999). However, lacking more precise information about these variables necessitates dealing with them as if they were random.

The question if treatments are effective, i.e. different treatments have a ‘substantially’ different influence on the target variable, is now transformed to the question if the variation of the target variable is mostly ‘explained’ by the variation of the treatment effect, or if it is due to other influential variables, or even if it is due only to variation of those parts which are modelled to be random (and which are not yet open to causative explanation). However, the split into causative and random parts, the split of causative parts into treatment and other explanatory effects is not unique and may influence of course the final findings heavily.

The question when the causal part of the model is big enough to be judged as relevant is a technical one met by several specific significance tests, which should not worry us here. It is only important to state that these procedures rely on an investigation how pure randomness would influence the target variable.

Two model situations to equalize other influential variables

Generalization of findings from samples to the population: To get reliable information about the mean of a population, a sample is taken from that population and the sample mean is taken as an estimate. If ‘measuring’ a single object of the population may be modelled merely by ‘natural’ variation of randomness, then all the properties of that randomness from the last section may be applied to an artificial summing or calculating the mean value of all the objects. Hence, the normal approximation and confidence intervals can be applied. This modelling is justified by a random sampling procedure to select objects for the sample to be ‘measured’. It is the random sampling which guarantees – with the exception of some calculable risk – that the results are generalizable, i.e. that the confidence interval covers the ‘true’

mean of the population. In this sense, with random sampling the sample drawn is *representative* for the population (see also Borovcnik 1992).

Thinking that samples are representative of populations if the sampling is ‘purely’ random is a key concept for the generalization of empirical findings. Randomness avoids all conceivable selective properties, which would lead to biased samples, or, randomness equalizes all selective properties so that finally the sample is representative for the population. Any arbitrary procedure to select the sample is prone to systematic, unforeseeable and uncontrollable deviations.

Generalization of differences between two samples: Often, a comparison between two (or more) groups has to be made. The groups have got a different treatment – one for example may be a special medication for insomnia, the other only a placebo (a harmless substitute without a pharmaceutical substance). ‘Is the medication more effective than the placebo?’ is the decisive question. Of course, the two groups have to be as equal as possible with respect to all characteristics that could influence the effect of the treatment: People have not to know that they get the placebo; people in the treatment group should not represent the most persistent cases already proven insensitive to any treatment, and so forth. Strictly speaking, should the groups to be drawn purely randomly from the population? With the exception of very few cases, this could not be fulfilled in practice.

However, if the investigated group as a whole does not differ substantially from the population, it is sufficient that a random *attribution* process establishes the subgroups for the different treatments, i.e. randomness decides to which group the next patient will belong. This random attribution should equalize all differences in the objects across the whole subgroups, which could have a causative influence on the target variable representing ‘success’ of the treatment. Insofar as the random attribution should eliminate, or better equalize all causative elements that could make the subgroups different (i.e. equalize the effect of all confounding variables), the actually observed differences then may be attributed solely to the treatment and are thus generalizable.

5 Statistical Thinking

A brief introduction into the debate is given against the background that at all times the argument was used that there is more to probability and statistics than is contained in the mathematical version of the pertinent concepts – a special kind of thinking was advocated; several authors refer to an outstanding role of an interplay between intuitions and formal concepts (see Fischbein 1975, or Borovcnik 1992).

The educational debate on probabilistic and statistical thinking

The educational debate on probabilistic and statistical thinking is a long-ongoing one. Even in times as early as the 70’s when the accent was heavily put on the mathematical and probabilistic part of the curriculum, a special type of thinking was argued to be behind the formal concepts – probabilistic *thinking*. It was not quite clear what could be understood by that but the argument was that there is some

additive not yet included in the mathematical concepts. However, when put to a crucial test, either mystique arguments were used to describe and justify what probabilistic reasoning is, or it was reduced to key ideas in the mathematical development of the concepts.

Heitele (1975), for example, gave a catalogue of fundamental ideas, which on the whole should constitute the various dimensions of probabilistic thinking. His list is reading like the titles of the chapters in a mathematical textbook on probability but could always also be attributed to some more general idea at second inspection:

- Calibrating the degree of confidence
- The probability space
- The rule of addition
- Independence
- Uniform distribution and symmetry
- Combinatorics to count equally likely cases
- Urn models and simulation,
- The idea of a sample to represent a population
- The idea of a random variable and its distribution
- Laws of large numbers.

There were a number of attempts to get a clearer image of the fundamental ideas behind probabilistic thinking. For example, Borovcnik (1997) endeavoured to arrange the ideas around the idea of information as a central hinge between individual's intuitions and the formal concepts of the mathematical theory:

- Probability as a special type of information about an uncertain 'issue'
- The idea of revising information when faced with new evidence
- To make transparent which information is used—also in the simulation of situations
- To condense information to a few numbers (thus eliminating randomness)
- To measure the precision of information
- To guarantee the representativity (=generalizability) of partial information
- To improve the precision of information

The roots of probabilistic reasoning in the psychological research go back to Kahneman and Tversky. They identify various tendencies in individual's behaviour to wrongly re-interpret problems and solve them in a way different to the accepted standard. Their approach of misconceptions had a great impact on further educational research. In an empirical investigation on children's behaviour, Green (1983) found abundant misconceptions that were to be expected according to Kahneman and Tversky. Scholz (1991) tried a constructive approach of a cognitive framework for individuals to allow for probabilistic reasoning in an adequate manner (i.e. to accept a standard interpretation of situations and end up with acceptable solutions). The work of Fischbein (1975) is devoted to develop instructive approaches towards developing a sound interplay between individual's intuitions and formal concepts of mathematics as a key to develop probabilistic reasoning.

From the 80's on, initiated by the EDA discussion started by Tukey (1977), the focus shifted towards statistical thinking and reasoning. The new motto then was numeracy, i.e. to learn to understand data and the information, which lies in them. The shift also involved more applications, real-life situations, all leading away from mathematics and probability – and also away from games of fortune. In the German debate at that time also, procedures of formal statistical inference entered the stage and won much attention in the reform of curricula. (The high goals of those times have now made place for more realistic ones, especially enabled by the simulation technique and the resampling idea for statistical inference.) Numeracy and statistical reasoning internationally were promoted by big projects in the USA, beginning with the Quantitative Literacy project, see for example Scheaffer 1991.

Numeracy and 'graphicacy' were targeted at simple but intelligent data analysis alongside the techniques of descriptive statistics and EDA. The role of the context where the data stem from, for the interpretation of results, received increasing attention; it became undisputed that a sound analysis of data and results is not really possible without reference to the context. In the German debate, Biehler (1994) is seeking a balance between probabilistic and statistically loaded curricula, as courses biased towards data analysis would cause an all-too simple and probability free conception of stochastics.

However, asked what statistical thinking could be, no one gave a clear answer. This unsatisfactory circumstance was the starting point for Wild and Pfannkuch to integrate ideas from empirical research.

Statistical thinking as contrasted to probabilistic thinking is involved in all steps from the provisional problem out of a context, across all cycles of making the problem more precise and model it, to a final model substantiated by the data. Statistical thinking is tightly related to the process of empirical research to filter out generalizable findings from empirical data. More or less, statistical thinking might be associated with strategies to increase knowledge.

The approach towards statistical thinking by Wild and Pfannkuch

Here, a brief discussion of statistical thinking along the lines of the approach by Wild and Pfannkuch (1999) will be given. According to their approach, four dimensions of that type of thinking should be regarded:

- The investigative cycle
- The interrogative cycle
- Types of thinking involved
- Dispositions

The investigative cycle is a systems analysis approach toward the initial (research) questions. It involves the components of

Problem → Plan → Data → Analysis → Conclusions,

which might be run through several times for refinement.

In the *problem* phase it comprises to grasp the “system dynamics” of describing the target variable, i.e. to find ‘all’ relevant explanatory variables, and to reflect about possible confounding variables. Also, assumptions of relations, hypotheses on relations from relevant other investigations or theories have to be used to clarify and “define problems”. Omissions in that phase give rise to many failures in empirical research.

In the *planning* phase, the investigator has to deal with the development of a “measurement system” for all the included variables. A “design” has to be developed on how the “sampling” and the “data management” is actually to be undertaken. A “pilot study” should give indications as to whether the plan is adequate and practicable.

The *data* phase comprises “data collection, data management and data cleaning”.

The *analysis* phase comprises “data exploration, planned analyses, unplanned analyses, and hypothesis generation”.

The *conclusions* phase consists of “interpretation, conclusions, new ideas, and communication”.

The interrogative cycle represents the strategic part of the investigation and includes the following phases (see Wild and Pfannkuch 1999 for details):

Generate → Seek → Interpret → Criticise → Judge

Types of thinking: Wild and Pfannkuch (1999) discern between general types of thinking and those fundamental to statistical thinking. For the general types, they list “strategic, seeking explanations, modelling, applying techniques”. Specific to statistical thinking they list

- Recognition of need for data
- Transnumeration
 - changing representations to engender understanding [...]
- Consideration of variation
 - noticing and acknowledging,
 - measuring and modelling for the purposes of prediction, explanation, or control
 - explaining and dealing with investigative strategies
- Reasoning with statistical models
- Integrating the statistical and the contextual
 - information, knowledge, conception

Dispositions amount to the psychological side of investigations and comprise the following attitudes: Scepticism, imagination, curiosity, openness (to ideas that challenge preconceptions), a propensity to seek deeper meaning, being logical, engagement, and perseverance.

Wild and Pfannkuch (1999) then proceed to the split of sources of variation in data into real (a characteristic of system) and induced (by data collection).

Finally they come up with the explanation of the regularities, the model found on basis of the data analysis process. They refer to statisticians who see the biggest contribution of their discipline in the isolation and modelling of ‘signal’ in the presence of ‘noise’. Then they discuss about the relative and interchangeable character of random and causal influences. They refer to randomness as “just a set of ideas, an abstract model, and a human invention which we use to model variation in which we can see no pattern”. In that they come close to the scenario character of probability described here.

6 Conclusion

Questions central to probabilistic and statistical thinking have been raised. They should clarify that both types of thinking are intermingled and are not easily described. From the discussion, however, crucial components of these types of thinking should become clearer. There will be no simple answer also after further endeavours into that topic. Even if the ideas are not easily described, the foregoing exposition and the examples illustrate how pertinent thinking is organized, to which end it could serve, and how such thinking is blurred.

The role and eminent importance of data, the context where they stem from, and the attitude of empirical research should become quite clear from the discussion. The interpretation of probability statements as scenario figures assisting in a broader problem solving process to come up with a more transparent decision may become more accepted by the examples outlined in the paper. The splitting of variation in data into causative and random parts in the search of explaining, predicting, and controlling phenomena may be a guideline for further attempts to clarify the issues of statistical thinking.

References

- Bea, W. and Scholz, R.: ‘Graphische Modelle bedingter Wahrscheinlichkeiten im empirisch-didaktischen Vergleich’, *Journal für Mathematik-Didaktik* 16, 299-327.
- Bentz, H. J.: 1983, ‘Zum Wahrscheinlichkeitsbegriff von Chr. Huygens’, *Didaktik der Mathematik* 11, 76-83.
- Biehler, R.: ‘Probabilistic Thinking, Statistical Reasoning, and the Search for Causes – Do We Need a Probabilistic Revolution after We Have Taught Data Analysis?’, in *Research Papers from the Fourth International Conference on Teaching Statistics*, Marrakech 1994, The International Study Group for Research on Learning Probability and Statistics, University of Minnesota [available from <http://www.mathematik.uni-kassel.de/~biehler>].
- Borovcnik, M.: 1992, *Stochastik im Wechselspiel von Intuitionen und Mathematik*, Bibliographisches Institut, Mannheim.

- Borovcnik, M. 1997, 'Fundamentale Ideen als Organisationsprinzip in der Mathematikdidaktik', *Didaktik-Reihe der Österreichischen Mathematischen Gesellschaft* 27, 17-32.
- Borovcnik, M. and Peard, R.: 1996, 'Probability', in A. Bishop e. a. (eds.), *International Handbook of Mathematics Education*, part I, Kluwer Academic Publishers, Dordrecht, 239-288.
- Falk, R. and Konold, C.: 1992, 'The Psychology of Learning Probability', in F. Sheldon and G. Sheldon, *Statistics for the Twenty-First Century*, MAA Notes 26, The Mathematical Association of America, 151-164.
- Fischbein, E.: 1975, *The Intuitive Sources of Probabilistic Thinking in Children*, D. Reidel, Dordrecht.
- Freudenthal, H.: 1980, 'Huygens' Foundation of Probability', *Historia Mathematica* 7 (2), 113-117.
- Gigerenzer, G., Hoffrage, U., and Ebert, A.: 1998, 'AIDS Counselling for Low-Risk Clients', *Aids Care* 10 (2), 197-211.
- Green, D. R.: 1983, 'A Survey of Probability Concepts in 3000 Pupils Aged 11-16 Years', in *Proceedings First International Conference on Teaching Statistics*, vol 2, Teaching Statistics Trust, 766-783.
- Heitele, D.: 1975, 'An Epistemological View on Fundamental Stochastic Ideas', *Educational Studies in Mathematics* 6, 187-205.
- Hoffrage, U., Lindsey, S., Hertwig, R., and Gigerenzer, G.: 2000, 'Communicating Statistical Information', *Science* 290, 2261-2262.
- Hoffrage, U., Gigerenzer, G., Krauss, S., and Martignon, L. (2002). 'Representation facilitates reasoning: What natural frequencies are and what they are not.', *Cognition* 84, 343-352.
- Huygens, C.: 1657, 'De ratiociniis in ludo aleae', in F. v. Schooten: *Exercitationes mathematicae*, Leyden.
- Kahneman, D. and Tversky, A.: 1972, 'Subjective Probability: A Judgement of Representativeness', *Cognitive Psychology*, 430-454.
- Kahneman, D., Slovic, P. and Tversky, A.: 1982, *Judgement under Uncertainty: Heuristics and Biases*, Cambridge Univ. Press, Cambridge.
- Kissane, B.: 1981, 'Activities in Inferential Statistics', in A. P. Shulte and J. R. Smart, *Teaching Statistics and Probability*, National Council of Teachers of Mathematics, Reston, Virginia, 182-193.
- Krauss, S., Martignon, L., Hoffrage, U., and Gigerenzer, G.: 2002, 'Bayesian Reasoning and Natural Frequencies: A Generalization to Complex Situations' (submitted for publication).

- Montgomery, D. C. : 1991, *Design and Analysis of Experiments*, J. Wiley & Sons, New York.
- Quetelet, A.: 1835, *Sur l'homme et le développement des ses facultés, ou Essai de physique sociale*, Paris.
- Riemer, W.: 1991, 'Das '1 durch Wurzel aus n'-Gesetz – Einführung in statistisches Denken auf der Sekundarstufe I, *Stochastik in der Schule* 11, 24-36.
- Scheaffer, R.: 1991, 'The ASA-NCTM Quantitative Literacy Project: An overview', in D. Vere-Jones (ed.), *Proceedings Third International Conference on Teaching Statistics*, International Statistical Institute, Voorburg, 45-49.
- Scholz, R.: 1991, 'Psychological Research in Probabilistic Understanding', in R. Kapadia and M. Borovcnik, *Chance Encounters: Probability in Education*, Kluwer Academic Publishers, Dordrecht, 213-254.
- Steinbring, H.: 1991, 'The Theoretical Nature of Probability in the Classroom, in R. Kapadia and M. Borovcnik, *Chance Encounters: Probability in Education*, Kluwer Academic Publishers, Dordrecht, 135-166.
- Stigler, S. M.: 1986, *The History of Statistics. The Measurement of Uncertainty before 1900*, Harvard Univ. Press, Cambridge, Mass.
- Tukey, J. W.: 1977, *Exploratory Data Analysis*, Addison Wesley, Reading.
- Vancsó, Ö. (ed.): 2003, *Matematika 10, 11*, Muszaki Kiado (Textbook in Hungarian).
- Vancsó, Ö.: 2004, 'Inverse probabilities in everyday situation (Bayesian-type problems)', paper distributed in the *TSG 11: Research and Development in the Teaching and Learning of Probability and Statistics* (L. Jun, J. M. Wisenbaker org.) at ICME 10, Kopenhagen.
- Wild, C. and Pfannkuch, M.: 1999, 'Statistical thinking in empirical enquiry', *International Statistical Review* 67, 223-265 (with discussion).

Papers on Statistics Education

Papers by:

Carmen Batanero

Marie El Nabbout

Vincent Lonjedo

Maria Meletiou-Mavrotheris,

Robert Peard

Anna Maria Serradó

A STRUCTURAL STUDY OF FUTURE TEACHERS' ATTITUDES TOWARDS STATISTICS

Carmen Batanero, *Universidad de Granada, Spain*

Assumpta Estrada, *Universidad de Lérida, Spain*

Carmen Díaz, *Universidad de Granada, Spain*

Jose M. Fortuny, *Universidad Autónoma de Barcelona, Spain*

Abstract: *We analyse the main components of 367 future teachers' attitudes towards statistics through their responses to the Survey of Attitudes Toward Statistics (SATS) scale. Analysis of components correlations and factor analysis serve to describe the relationships among main components of teachers' attitudes and compare with previous research. Our results suggest that the four components (difficulty, value, affective and cognitive factors) described by the SATS authors might not appear clearly separated in future teachers. Relationships with other variable are also explored.*

1. Introduction

Statistics is increasingly taking part in the primary school mathematics curriculum; yet most primary school teachers have little experience with statistics and share with their students a variety of statistical misconceptions and errors (Stohl, 2005). An additional factor that affects teaching performance is teachers' attitudes towards the topic.

Background

In conceptualising the mathematics education affective domain, McLeod (1992) distinguishes among emotions, attitudes and beliefs. Attitudes are intensive feelings, relatively stable, which are consequence of positive or negative experiences over time in learning a topic (in this case statistics). The interest towards beliefs, attitudes, and expectations that students bring into statistics classrooms is increasing in statistics education, since "*such factors can impede learning of statistics, or hinder the extent to which students will develop useful statistical intuitions and apply what they have learned outside the classroom*" (Gal & Ginsburg, 1994, p. 1). In educating the teachers we should follow the advice by these authors, and make statistics teaching enjoyable and useful for them. In this way, the teachers will develop an appreciation for how the application of statistics is useful in their professional and personal lives and for their students.

In the past two decades a large number of instruments to measure attitudes and anxiety toward statistics have been developed in order to assess the influence of emotional factor on students' training (see Carmona, 2004 for a review). Research on students' attitudes towards statistics is increasing; although it is still scarce when compared with research related to attitudes towards mathematics and has not

focussed specifically on teachers. Multidimensional studies of attitudes are more frequent in the past years and try to establish the basic elements that conform them. This research is summarised in Gal & Ginsburg (1994) and more recently in Carmona (2004), who also analyses the psychometrical features of the different instruments. In this paper we complement our previous studies (Estrada, 2002; Estrada, Batanero & Fortuny, 2003, 2004) of teachers' attitudes towards statistics. The aim is describing the structure of these attitudes and its relationships with statistical knowledge, number of statistics courses taken, speciality and gender, as a base to develop formative actions directed to these teachers.

2. Method

Participants were 367 pre-service teachers training in different specialities at the Faculty of Education, Lérida, Spain. The survey was given to the students as a part of the mathematics course and before they were taught the statistics unit. Student attitudes were measured using the Survey of Attitudes toward Statistics (SATS) (Schau, Stevens, Dauphine, & Del Vecchio, 1995). The authors define attitude towards statistics as a multidimensional construct, composed of different analysable dimensions (p.57), which are structured in four components. The SATS (included in appendix) is a 28-item Likert instrument that has four sub-scales:

- *Affect*: Positive or negative feelings concerning statistics: items 1, 2, 11, 14, 15, 21.
- *Cognitive competence*: Perception of self-competence, knowledge and intellectual skills when applied to statistics: items 3, 9, 20, 23, 24, 27.
- *Value*: Appreciation of statistics usefulness, relevance and value of statistics in personal and professional life: items 5, 7, 8, 10, 12, 13, 16, 19, 25.
- *Difficulty*: Perceived difficulty of statistics, as a subject: items 4, 6, 17, 18, 22, 26, 28.

In our research, each statement was valued in a range from 1 to 5, where 1 indicates "Strongly Disagree" and 5 indicates "Strongly Agree". According to Gal, Ginsburg and Schau (1997), *scores on affect and cognitive competence scales are strongly related to each other. Scores on value and difficulty are moderately related to those on affect and cognitive competence, but not related to each other* (pg. 44). The reason for choosing this scale is that it allows us to analyse the structure of responses and its reliability and validity have been assessed through different research. In our sample the value of the coefficient alpha was 0,89 for the total score and 0,80, 0,73, 0,77 and 0,7 for the affective, cognitive, value and difficulty factors, respectively.

Knowledge of statistics in future teachers was assessed with items 1, 2, 3, 4, 7, 12, 15, 16 and 17 of the SRA test, with a total of 19 sub items (Garfield, 2003). Each item describes a statistics problem and offers several choices of responses, both correct and incorrect in a multiple-choice format. Different alternatives include a statement of reasoning, explaining the rationale for a particular choice. Given the

limitation of space we do not include the particular items (the whole instrument is included as an Appendix in Garfield's paper that can be downloaded from the SERJ web page). The particular items we used assess knowledge of main content in the statistics curriculum for primary school in Spain.

The following types of reasoning were included in the selected SRA items: reasoning about data, interpreting graphs, reasoning about average and spread, uncertainty and bias in sampling, and association. In addition to determining types of reasoning skills, these items also identify the following misconceptions or errors in reasoning: misconceptions involving averages, outcome approach, confusing correlation with causality, law of small numbers, and representativeness heuristics (Garfield, 1998).

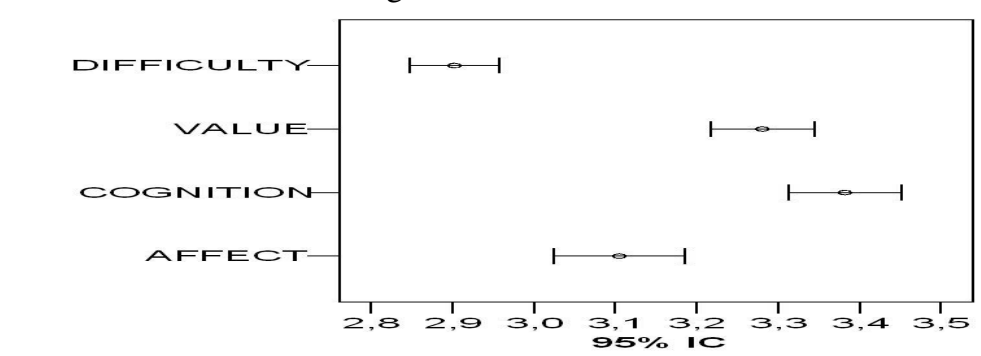
3. Results and discussion

Below we present three types of results: a) future teachers' attitudes towards statistics and its components; b) future teachers' difficulties in some elementary statistics concepts; and c) effect of statistical knowledge, previous study of statistics, gender and speciality on attitudes.

3.1. Attitudes towards statistics

In Figure 1 we compare the average score per item in each of the four components (dividing the total score in the component by the number of items, in order to get a homogeneous scale). Negative statements were reverse coded. Since a score of 3 corresponds to the indifference point, our results suggest that participants saw statistics as slightly difficult (score under theoretical mean), had slight positive valuation of the topic and positive perception of their own capacity to learn it, and were a little positive in their feelings towards statistics.

Figure 1. 95% Confidence intervals



On the other hand the study of correlation among components (Table 1) suggest the order of impact of these components on the global attitude, which is scarce in the case of difficulty. That is, future teachers' attitudes were little influenced by considering the topic either difficult or easy. Contrary to Gal, Ginsburg and Schau' (1997) results, affect is highly correlated with cognition; and difficulty and value have small correlation. Therefore teachers seem to value statistics regardless of the perceived difficulty and feelings towards the topic that seem to depend on the perceived self capacity for learning.

Factor analysis

We carried out an exploratory factor analysis; this is a technique used in previous studies of attitude towards statistics (e.g. Mastracci, 2000). Our data fulfilled the assumptions to apply the method (more than 10 cases per variable, experimental unit, factorizability of the correlations matrix, normality, linearity and lack of multicollinearity). The matrix determinant, Barlett test of sphericity and Kaiser - Meyer - Olkin index all gave values in the appropriate range.

The method of initial factor extraction was principal components, also used by Mastracci's (2000) in his research with undergraduates. This method does not distort the data, since it only involves a change of reference in the variables vector space. Following Cuadras' (1991) recommendations and the general principle of interpretability we decided to retain 5 factors. The total variance and percent of variance explained by each factor are displayed in Table 2. The greater weight of the first factor and the similar relevance of the remaining factors are visible.

Table 1. Pearson' correlation coefficients

Component	Affect	Cognitive	Difficulty	Value
Total score	0,88	0,87	0,75	0,77
Affect	1,00	0,78	0,47	0,64
Cognitive		1,00	0,45	0,63
Difficulty			1,00	0,33
Value				1,000

Table 2. Factor analysis summary

Factor	Total	% Variance	Cumulative %
1	7,36	26,28	26,28
2	2,22	7,94	34,21
3	1,68	6,00	40,21
4	1,26	4,47	44,71
5	1,20	4,27	48,98

To facilitate the interpretation of the retained factors we rotated the axes using the Varimax method, an orthogonal rotation that maximizes variance and does not distort the data. In table 3 we present the rotated factorial scores. In order to facilitate interpretation, the variables appear in decreasing order according to their contributions to the first factor. We include a sentence to remember the item content (expressed as a positive attitude) and a letter to describe the component to which the item belongs (A=affect, C= Cognitive competence; V=Value; D= Difficulty). Below we interpret the factors.

- *First factor: affective and cognitive components.* This factor explains 26.2% of the total variance and includes most items in the affective and cognitive components. Variables in these two domains are matched in pairs (I understand: I feel secure; I have ideas: lack of frustration, etc.). In our sample these two components are related, contradicting the opinion of Gal, Ginsburg and Schau (1997). These results suggest the extent to which the teachers' affection towards statistics might be conditioned by their understanding of the topic. Of course this might be a specific characteristic of teachers, but reinforces our view that the statistical training of teachers should be increased, since a teacher who feels insecure or scared about a topic is unlikely to support its teaching.

Table 3. Varimax rotated factors. Principal components extraction

Statement	Component	Item	Factor loadings				
			1	2	3	4	5
Security	A	I2	0,76				
Understanding	C	I3	0,72				
Lack of frustration	A	I11	0,58		0,35		
Having ideas	C	I9	0,57	0,32			
Not many errors	C	I20	0,51		0,35	0,41	
No fear	A	I21	0,50		0,43		
Concepts easy to understand	C	I27	0,50		0,39		
Easy	D	I6	0,45		0,42	0,37	
Lack of Stress	A	I14	0,42		0,31		
Applicability	V	I12		0,71			
Useful	V	I10		0,71			
Wide presence	V	I16		0,71			
Professional value	V	I19		0,68			
Worth	V	I5		0,46			
Relevance	V	I25	0,37	0,44			0,34
Frequent use	V	I13		0,36			
Understand equations	C	I24			0,74		
Easy formulas	C	I4	0,39		0,65		
I can learn	C	I23			0,64		
I like	A	I1	0,46		0,53		
I enjoy	A	I15	0,37		0,49	0,35	
Most people learn	D	I17			0,30		
Lots of computation	D	I22				0,77	
Very technical	D	I26				0,71	
New way of thinking	D	I28					-0,72
Increases employability	V	I8		0,44			0,55
Should be a requirement	V	I7		0,37	0,39		0,55
Requires discipline	D	I18				0,33	-0,45

- *Second factor: Value.* This factor groups all those items that present statistics as an important tool in different domains. It represents teachers' beliefs on the relevance of statistics in the society, their own training and the school curriculum. These items have little influence on the other factors, so that our results agree with those of Mastracci (2000), who have also got value as a separated component of the attitudes.
- *Third factor: affective and cognitive components.* The items in factor 1 are repeated again with slight variations.
- *Fourth and fifth factor: difficulty.* These factors group most statements related to sources of difficulty in the study of statistics. Here we again agree with Mastracci (2000) who obtained difficulty as a separate component. The relative strong

weight in items 22 and 26 suggests that many students might associate the discipline's difficulty to the mathematical technical features. The negative correlations of some difficulty items connected to value on factor 5 suggest that professional relevance of statistics might be perceived in inverse relation to the degree of difficulty of the matter.

3.2. Statistical knowledge

Table 4. Percent of correct responses to SRA sub items

Item	Item in SRA	% Correct	Item content
1	1	46.9	Mean as best estimator in presence of outliers
2	2	76	Interpreting probability
3	3	59.5	Outcome approach
4	4	72.2	Mean as representative value
5	15	73.3	Comparing two groups (graphs)
6a	16a	84.2	Sample size
6b	16b	49.3	Correlation vs causality
6c	16c	42.8	Correlation vs causality
6d	16d	69.5	Sample size
6e	16e	51.8	Correlation vs causality
7	17	33.2	Relating mean to total
8	12	73.3	Sampling Variability as related to sample size
9a	7a	77	Sample mean as estimator of sample population
9b	7b	74.9	Adequate sample size
9c	7c	72.2	Adequate sample size
9d	7d	58.3	Bias in sampling
9e	7e	69.1	Random vs conglomerate sampling
9f	7f	70.6	Bias in sampling
9g	7g	84.5	Estimation in random sampling

In Table 4 we present the percentage of correct response to each SRA item. In case an item is composed by several sub items results are presented for each sub item. Even when the difficulty was low or moderate in most items, results suggest that an important percent in the sample of future teachers did not understand some elementary statistical concepts they have to teach their future students correctly.

For example 45% of participants did not take into account outliers when computing averages; 27.8% of them showed the outcome approach, around 45% of them confused correlation with causality in different questions; 23.8 % did not relate the mean with the total, more than 30% was insensible to sample bias in different items, 15% considered estimation was not possible because of random fluctuation, 30% did not understand the idea of conglomerated sampling and around 30% have other errors related to sampling.

3.3. Factors affecting future teachers' attitudes

In table 5 we present Pearsons' correlation coefficients between total score in the items taken from the SRA questionnaire (these score ranged between 0 and 19, since there was a total of 19 sub items) and the attitudes total and component scores. The only non significant component was difficulty, which suggests that participants considered statistics to be difficult regardless of their knowledge. Positive small correlations in the other components suggest that attitude and its components in general tend to improve a little with increased knowledge.

Table 5. Correlation coefficients

Correlation with SRA score	
Total score	0,23*
Affect	0,20*
Cognitive	0,26*
Difficulty	0,09
Value	0,22*

Table 6. Results from Variance Analysis

Source	d.f.	F	P value
Courses taken or not	1	10,10	0
Gender	1	3,26	0,07
Speciality	5	1,84	0,11
Interaction	5	0,60	0,7

* pvalue < 0.01

In table 6 we present results from Variance analysis of total score in the attitudes scale as regards different factors. The only significant factor was the number of statistics courses previously taken (in secondary school) by the participants (this number ranged between 0 and 3). Detailed analyses of scores showed that attitudes improved consistently with this number. Similar analyses showed the improvement with the number of courses in all the components (improving systematically) except by difficulty.

4. Conclusions

Our results suggest that the four components (difficulty, value, affective and cognitive factors) described by the SATS authors might not appear clearly separated in future teachers, although we coincided with Mastracci (2000) in obtaining separate components of difficulty and value. The affective and cognitive components are linked in the first and third factor, which indicates the extent to which affect is influenced by the understanding of the matter in our sample. This conjecture is reinforced in the fact that the number of previous courses of statistics was the only significant factor affecting teachers' attitudes and by positive correlations with total score in the knowledge test.

There is general consensus in the mathematics education community that teachers need a deep and meaningful understanding of any mathematical content they teach. This type of understanding was not present in our study as regards very elementary statistical concepts. Sullivan (2003) suggests that the issues of mathematical content knowledge and beliefs about the nature of mathematics are formed by experiences prior to the teacher education program. Consequently it would be useful for teachers trainers to consider the appropriate formative experiences that will foster the

prospective teachers' capacity for ongoing statistical learning, help them reflect on the nature of statistics, and help them value statistics knowledge and literacy in improving the education of all the citizens.

Acknowledgement: Research supported by the grant SEJ2004-00789, MEC-FEDER, Spain.

References

- Carmona, J. (2004). Una revisión de las evidencias de fiabilidad y validez de los cuestionarios de actitudes y ansiedad hacia la estadística. *Statistics Education Research Journal*, 3(1), 5-28. On line. Available from: [http://www.stat.auckland.ac.nz/~iase/serj/SERJ3\(1\)_marquez.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ3(1)_marquez.pdf)
- Cuadras, C.M. (1991). *Métodos de análisis multivariante*. Barcelona: Eunibar.
- Estrada, A. (2002). *Análisis de las actitudes y conocimientos estadísticos elementales en la formación del profesorado* (Análisis of attitudes and elementary statistical knowledge in educating the teachers). Ph.D. Universidad Autónoma de Barcelona.
- Estrada, A., Batanero, C & Fortuny, J. M. (2003) Actitudes y estadística en profesores en formación y en ejercicio (Attitudes and statistics in service vs. prospective teachers). *Proceedings of 27 Spanish National Conference of Statistics and Operational Research. Universit of Lleida. CD ROM*
- Estrada, A., Batanero, C & Fortuny, J. M. (2004). Un estudio comparado de las actitudes hacia la estadística en profesores en formación y en ejercicio (A comparative study of attitudes towards statistics in teachers). *Enseñanza de las ciencias*, 22(2), 263-274.
- Gal, I. & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: towards an assessment framework. *Journal of Statistics Education*, 2(2). On line. Available from: <http://www.amstat.org/publications/jse/v2n2/gal.html>
- Gal, I., Ginsburg, L., & Garfield, J. B. (1997). Monitoring attitudes and beliefs in statistics education. In: I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 37-51). IOS, Press, Voorburg.
- Garfield, J. B. (1998). The statistical reasoning assessment: Development and validation of a research tool. In L. Pereira- Mendoza (Ed.), *Proceedings of the 5th International Conference on Teaching Statistics* (vol. 2, 781-786). Singapore: International Statistical Institute.
- Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, 2(1), 22-38. On line. Available from: [http://www.stat.auckland.ac.nz/~iase/serj/SERJ2\(1\).pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ2(1).pdf)
- Mastracci, M. (2000). *Gli aspetti emotivi nell'evoluzione dell'apprendimento della statistica e della sua valutazione. Un caso di studio sugli studenti di SSA*. Tesi di Laurea. Università La Sapienza di Roma.

- McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grows (Eds.), *Handbook of Research on Mathematics Teaching and Learning*, 575-596. Macmillan N.C.T.M. New York.
- Schau, C., Stevens, J., Dauphine, T. & Del Vecchio, A. (1995). The development and validation of the survey of attitudes towards statistics. *Educational and Psychological Measurement*, 55 (5), 868-875.
- Stohl, H. (2005). Probability in teacher education and development. In G. Jones (Ed.). *Exploring probability in schools: Challenges for teaching and learning*, 345-366. New York: Springer.
- Sullivan, P. (2003). Editorial. Incorporating knowledge of and beliefs about mathematics into teacher education. *Journal of Mathematics Teacher Education*, 6, 293-296.

Appendix. Sats scale

(Note. Each item below should be followed by a 5-point scale, ranging from 1 (strongly disagree) to 5 (strongly agree).

1. I like statistics.
2. I feel insecure when I have to do statistics problems.
3. I have trouble understanding statistics because of how I think.
4. Statistics formulas are easy to understand.
5. Statistics is worthless.
6. Statistics is a complicated subject.
7. Statistics should be a required part of my professional training.
8. Statistical skills will make me more employable.
9. I have no idea of what's going on with statistics.
10. Statistics is not useful to the typical professional.
11. I get frustrated going over statistics tests in class.
12. Statistical thinking is not applicable in my life outside my job.
13. I use statistics in my everyday life.
14. I am under stress during statistics class.
15. I enjoy taking statistics courses.
16. Statistics conclusions are rarely presented in every day life.
17. Statistics is a subject quickly learned by most people.
18. Learning statistics requires a great deal of discipline.

Working Group 5

19. I will have no applications for statistics in my profession.
20. I make a lot of maths errors in statistics.
21. Statistics scares me.
22. Statistics involves massive computation.
23. I can learn statistics.
24. I understand statistics equations.
25. Statistics is irrelevant in my life.
26. Statistics is highly technical.
27. I find it difficult to understand statistical concepts.
28. Most people have to learn a new way of thinking to do statistics.

TEACHERS' REPRESENTATIONS OF INDEPENDENT EVENTS: WHAT MIGHT AN ATTEMPT TO MAKE SENSE HIDE?

Sylvette Maury, *Université René Descartes – Paris 5, France*

Marie Nabbout, *Université René Descartes – Paris 5, France*

Abstract: *Teachers tend to attribute meanings to certain mathematical concepts in order to make them more accessible to students. But to what might that lead, when teachers tend to attribute a common sense definition for independent events valid in the cases of “Chronological events” to other purely theoretical cases like “Stochastic independent events”? In this paper, we tried to study teachers’ representations of independent events based on their evaluation of students’ pre-prepared answers and on their answers to some direct questions. We mainly focused on studying the link that may exist between the common sense definitions that teachers attribute to “independency” and the confusion between independency and incompatibility, which is students’ most common confusion.*

1. Introduction

Steinbring (1986) distinguished between two cases of independent events that are not contradictory but are in opposition. In the first case, we talk about “Intuitive independency” where two events are said to be independent when “they are not influenced one by the other”, that is when they are associated to experiences occurring successively and where the independency is postulated. Such events are as well called “Chronologically independent events” or, “a priori independent events” (*cf.* Maury 1985) when the situation consists of successive “independent” events in the naïve sense of the term.

The second case of independency is the “formal independency” or the “stochastic independency” of events, which is based on the formal mathematical definition: “2 events A and B are independent $\Leftrightarrow P(A \cap B) = P(A) \times P(B)$ ”. In this case, no reference is made to any experiments or to any chronology, and the independency is only defined by the mathematical formula.

According to Maury (1984), the a priori independency does not present any necessity in the probability theory; but most teachers and textbooks’ authors do refer to it, while trying to attribute meaning to the mathematical concept of independency.

Sanchez (2000) observed in his study that teachers display the same confusions as students, and he attributed that to the fusion between the intuitive idea of independency and the abstract mathematical concept of independent events.

2. General Framework

The data on which our work is based are collected from a research¹ that aims at studying some representations of Lebanese (French speaking) high school mathematics teachers in probability and statistics and their methods of teaching.

¹ This paper is related to a doctoral dissertation carried by Marie Nabbout under the supervision of Sylvette Maury at Paris 5 – René Descartes University. The research consists of 4 individual interviews with 16 Lebanese mathematics teachers.

In this paper, we studied teachers' representations of independent events by analyzing, on one hand, their assessment of students' answers, and on the other hand, their answers to some direct questions.

First, we studied the comments and the arguments of teachers while assessing students' work on two problems (*cf.* Appendix 1) both about stochastic independent events². The work was done during (registered) individual interviews, where teachers were asked to assess (correct, grade and evaluate) four answers. Each answer was chosen with a specific intent, as explained in the following paragraph. Second, we studied teachers' answers to some direct specific questions³ related to independent events (*cf.* Appendix 2).

3. Choices of the tasks

As stated before, our concern in this paper was to study teachers' representations for independency, and accordingly we chose the problems and the corresponding answers.

The first point we intended to check was the recognition by teachers of stochastic independency situations, in other terms whether they are able or not to distinguish between stochastic and chronological independency. To do that, we chose the erroneous answers S4E3⁴, S6E2, and S6E3 (*cf.* Appendix 1) which are based on invalid "paraphrases" of the mathematical definition of independent events, but which correspond to the naïve meaning of "influence" that works in the case of chronologically independent events.

In fact, a valid paraphrase of the mathematical definition of independent events is '*The occurrence of A has no influence (or does not affect) the probability of B, thus A and B are independent*'. But, due to language imprecision, such a paraphrase becomes erroneous like P1: '*The occurrence of A has no influence (or does not affect) B, thus A and B are independent*'; or like P2: '*If A occurs, B may occur or not and if A does not occur, B may occur or not, thus A and B are independent*'. In P1 and P2, no references are made to any probability and hence these two statements are not valid in the case of stochastic independency.

Dupuis and Rousset-Bert (1998) observed such answers among students and they inferred that such answers indicate that students consider independency as an intrinsic property of the events, without taking into consideration their probability.

The other three erroneous answers, S4E1, S4E2, and S6E1 (*cf.* Appendix 1) are about the confusion between incompatible events and independent events, which is recognized as students' most common confusion. What we needed to check at this level was how teachers would evaluate these answers.

The last two answers S4E4 and S6E4 were correct.

² This is part of the work done during the 3rd interview with teachers, where teachers had to evaluate students' work on 8 different problems and where, for each problem, 4 answers were proposed. In this paper, we studied 2 problems S4 and S6, related to the stochastic independency. The other 6 problems were about chronological independency, conditional probability and simple probability.

³ This is part of the work done during the fourth interview with teachers.

⁴ In SiEj, i designates the number of the problem and j that of the answer. Thus, S4E3 stands for the third answer given to problem 4.

The following table summarizes the differences and similarities between the chosen answers:

The answer we chose hide misconceptions that we	Answer	Alignment with the correct answer
S4E1 – Q1	Confusion: Independency – Incompatibility	Aligned with the correct answer
	Confusion: Independency - Incompatibility	Different from the correct answer
S4E2 – Q1	Confusion: Independency – Incompatibility	Aligned with the correct answer
	Confusion: Independency - Incompatibility	Different from the correct answer
S4E3 – Q1	Paraphrase of the definition.	Different from the correct answer
	Paraphrase of the definition.	Aligned with the correct answer
S4E4	Correct answer	
S6E1	Confusion: Independency – Incompatibility	Different from the correct answer
S6E2	Paraphrase of the definition.	Aligned with the correct answer
S6E3	Paraphrase of the definition.	Aligned with the correct answer
S6E4	Correct answer	

needed to check whether teachers would recognize, and in case they would not be able to identify them, to see how they would defend their validity. The correct answers, which helped teachers in certain cases to regulate their judgments through comparison and to change their evaluation, were chosen on purpose because we wanted to observe the whole process that might occur in case of conflict; especially in problem 4 which consists of 2 questions, one being aligned with the expected answer, and the other not being aligned.

It is as well important to distinguish between problem 4 and problem 6. Even though problem 6, is a stochastic independency situation, still it may be interpreted in a “special” intuitive way and that by considering it as the cross product of 2 independent criteria, but that does not figure clearly in the proposed answers. Yet, there is no way to interpret problem 4 in a similar way.

As for the direct questions, the first served to classify and to compare teachers’ definitions of independent events; the second, to investigate the meaning that teachers attribute to the “independency” while discussing the equivalence (or not) of “non independent events” to “dependent events”.

Finally, it is important to mention that the repetition of certain questions at various moments during the experiment and in different forms was carried out in order to keep track of the persistence of certain ideas or their evolution⁵.

4. Results

In the following paragraphs, we will describe teachers’ behavior while evaluating the erroneous answers, and will discuss how we drew conclusions about their

⁵ The fourth interview took place at least 6 months after the third interview.

representations of independent events. These results are summarized in the following chart.

Teachers' evaluation of erroneous answers

NON DISTINCTION between stochastic and chronologic situations													DISTINCTION between stoch. and chrono. situations		
T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16
NON RECOGNITION of the confusion: Incompatible - Independent								RECOGNITION of the confusion: Incompatible - Independent							
T1	T2	T3	T4	T5	T6	T7	T14	T8	T9	T10	T11	T12	T13	T15	T16
					M2	M2	C.C	M3	M4						

M1: Teacher accepts S4E1 and S4E2 but recognizes the confusion later on.
M2: Teacher finds the method used in S4E1 and S4E2 sound but doesn't work.
M3: Teacher accepts S6E2 and S6E3.
M4: Teacher considers that common elements between A and B implies there is an influence of A on B.
CC: Cautious, careful attitude.

Teachers' representations of independent events

EXPLICIT CONFUSION: Incompatibility - Independency		IMPLICIT CONFUSION: Incompatibility - Independency						No Confusion			
T1	T2	T3	T4	T6	T7	T8	T12	T13	T14	T15	T16

4.1. Recognition of correct answers:

All teachers recognized with no trouble the correct answer in both situations.

4.2. Chronological independency and stochastic independency:

Most teachers did not distinguish between chronological independency and stochastic independency. In fact, there were 3 teachers out of the 16 who were implicitly able to recognize in I3⁶, that S4 and S6 are situations of stochastic independency, and only one of them distinguished between the 2 cases while defining independent events in I4. Few teachers acknowledged that there are cases where one cannot use common sense to interpret "independency". They considered these situations as complex or fuzzy, and they preferred to refer, in such cases, to the formula. But they failed to describe these cases and generally failed to distinguish between them before hand, which is the first difficulty observed with the teachers concerning this concept.

4.3. "Explicit"⁷ confusion between Incompatible events and Independent events:

The confusion between incompatible events and independent events is a very common error observed among students. This confusion was observed with four teachers in I3. Two of them accepted answers S4E1 and S4E2, and realized later on by comparison to S4E4 that the answers were wrong; but they were not able to refute them. The other two were later able to recognize the mistake and recognized the underlying confusion. What we can tell at this stage, about the last two teachers, is that their notion of independency seems to be inconsistent and it might hide a

⁶ From now on, we will use I3 to designate the third interview, and I4 for the fourth interview.

⁷ We used the term explicit to distinguish this case from another one used later in this paper, where we talk about 'implicit confusion between incompatible events and independent events'.

misconception (M1⁸). As for the first two, we can confirm the presence of the explicit confusion.

Definitely, this confusion was not observed in S6E1, which took place slightly after the evaluation of S4 and whose answer, though based on the same misconception, is wrong.

4.4. Recognition of the confusion between Independent events and Incompatible events:

It is true that the confusion between incompatible events and independent events only occurred with two teachers, while correcting S4E1 and S4E2, and that the other fourteen teachers refuted these answers. But, not all teachers who rejected these answers were able to justify why these answers were wrong, nor to recognize the underlying misconception.

As stated before, two of these 14 teachers had accepted S4E1 and S4E2 but realized later on the underlying confusion and accordingly refuted the answers (M1). One of them recognized the confusion on S6E1 but the other did not explain why he refused it.

Eight teachers had directly recognized the confusion in S4E2, six of them had recognized it in S4E1, and another six in S6E1. Only five teachers had recognized this confusion in the three cases. This might be due to the fact that S4E2 is the only answer in which the misconception is very explicit. S4E1 hides the same misconception, but it is said implicitly in sentences. As for S6E1, which is not aligned with the correct answer, it may be that the teachers did not need to reflect on the error, or it might equally be due to the fact that it resembles S4E1 more, since it is a sentence.

On the other hand, the eight teachers who did not accept S6E1 did not or could not explain why it was wrong. Three of them did the same for S4E1 and two of them for S4E2. There was only one teacher who couldn't tell why S4E1 is wrong, but was able to identify the misconception in the other cases.

What is interesting in this comparison is not the number of teachers who did not explicitly name this misconception, or could not identify it, but the significance that this non-recognition might hide.

Two teachers refuted the answer since it was not aligned with the correct one in both questions (S4) and explained that the reasoning on which the answer is based does not hold in both cases. This drives us to believe that these teachers found something credible in the given answers but were obliged to refute the answer since it is not aligned with the correct one. We assume the presence of a misconception with these two teachers (M2).

4.5. Results of S4E3, S6E2 and S6E3:

As we explained in paragraph 3, these three answers are wrong.

a. As stated before, three teachers had implicitly recognized these situations as ones of stochastic independency and easily refuted these three answers.

⁸ We labeled potential misconceptions by Mi.

b. Four teachers accepted part of the answer S4E3 but accepted S6E2 and S6E3. The explicit confusion between incompatibility and independence was observed in the case of one teacher, while M2 was assumed to be present with the other teacher. We assume the presence of a misconception (M3) with the other two teachers.

c. Three teachers did not accept S4E3 because “there is an influence of A on B”, and justified that by talking about common elements between A and B. Even though they considered that the answer shows an understanding of the definition of independent events, two of them refused S6E2 for the same reason, and were hesitant while trying to justify what they considered correct in S6E3. Misconceptions M1 and M3 were assumed to be present with two of them respectively. We assume the presence of a misconception M4 with the third teacher.

d. Two teachers did not accept these three answers without explaining why. They simply stated that it was wrong. One of them already showed the explicit confusion between independent events and incompatible events.

e. Five teachers were a bit hesitant in their justification. They did not accept the three answers. In S4E3, three of them stated that P1 makes sense. In S6E2, two of them stated the same thing and two others accepted the answer but found that it required justification. In S6E3, they all refused the answer but did not (or could not) justify why it was wrong. Misconceptions M1 and M2 were respectively assumed to be present with two teachers.

We described the four teachers who did not show any misconception (in d. and in e.) as teachers with a “cautious, careful” attitude.

4.6. Definition of Independent events (I4):

All teachers use mathematical formulas to define independent events, at some point during their teaching, since this is related to their math course. But, what we tried to check in this question is what accompanies or precedes these formulas, and if teachers distinguish between chronological independency and stochastic independency.

- One teacher, as stated in 4.1, explicitly distinguished between chronological independency and stochastic independency.
- Four teachers stated that they only defined “independent events” using formulas, since it is “more secure for the students”. Misconception M2 was assumed to be present with two of them. The third teacher belongs to the group who has the “cautious, careful” attitude.
- Nine teachers defined independent events as the paraphrase P1. Some of them even added to it other elements. Four of them stated that they would rather use a formula because it is safer. Misconceptions M3 and M4 were assumed to be present in two of them and the explicit confusion was observed in the case of the third teacher. The fourth teacher belongs to the group having the “cautious, careful” attitude.
- Two teachers defined independent events as a correct paraphrase of the mathematical intuitive definition, where they talked about the probability of the second event. M1 was assumed to be present with one of them.

- One teacher tried to explain his idea with an example of independent events, but ended up with an example of incompatible events; he was in the group of teachers with “cautious, careful” attitude. We assume the presence of the misconception M4 with him.

4.7. Non-independent events – Dependent events (I4):

- Six teachers insisted on using “non-independent events” and refused to use “dependent events”. Four of them explained that “non independent events” are not necessarily “dependent events”. One of these four had the “cautious, careful” attitude. Misconceptions M2 and M3 were already assumed to be present with the other teachers.
- Seven teachers found both terminologies similar.
- Three teachers preferred to use “dependent” rather than “non independent” since “dependent” is more meaningful. Misconception M1 was already assumed to be present with two of them.

5. Discussion: Implicit confusion between incompatible events and independent events.

In 4.5.b., we assumed the presence of misconception M3 with two teachers who accepted S6E3 and partially accepted S4E3 and S6E2. In fact, we think that the non-alignment of the 2 parts of S4E3 is what made them partially accept the answer.

In 4.5.c., we saw that three teachers considered that there is an “influence” of A on B since “there are common elements between A and B”. Misconceptions M1, M3 and M4 were assumed to be present with them.

In 4.6, we assumed the presence of M4 with one teacher.

Five of these teachers are among the nine teachers who defined in 4.6 independent events as a paraphrase based on “no influence of one event on the other”. The sixth uses only a formula since it is more secure.

Three teachers are among the ten who admitted in 4.7 that the terms “non independent events” and “dependent events” were alike. The other three refused to use “dependent events” as the opposite to “independent events” and two of them argued that non independent events may not be dependent.

We tried to hypothesize the confusion underlying the four misconceptions (M1, M2, M3 and M4) shared by these six teachers, by constructing a chain of terms they used during their evaluation or while defining or talking about independent events.

When teachers tried to make sense of the independency of events, they tended to use paraphrases of the definition, and they considered that:

- If there is no influence of one event on the other, then the two events are independent.
- If there is an influence of one event on the other, then the two events are non-independent.
- If there is a common element between A and B then there is an influence of A on B.
- Non-independent events are dependent events.
- “Dependent” means there is a relation or a bond.

We put these terms in a sequence that may reveal the succession of ideas for teachers. The chain of words thus obtained shows that the hidden misconception is simply the common confusion between incompatible events and independent events. In fact, the sequence is: Common element - Relation or Bond - Influence - Dependent events – Non-independent events. It can be presented in a simpler form: ***Common element – Influence – Non-independent events***.

In other words, if there is a common element between A and B, then there is an influence of A on B, and so A and B are not independent. Hence, non incompatible would lead to non independent. Following the same logic: Independent events \Rightarrow Incompatible events. We named this type of misconception: the implicit confusion between incompatibility and independence.

This confusion was directly observed with the two teachers who displayed M4. We assume that the same confusion is present with the four other teachers who had displayed M2 and M3.

6. Conclusion.

Teachers try to interpret the meaning of “independent events” because they do not want to teach this concept in a very abstract way, but want to make it accessible to their students. In fact, most teachers insist on the importance of making students understand “the real meaning” of this concept in order to overcome obstacles. However, this attempt to make a formal mathematical concept meaningful by attributing to it meanings that do not hold in all cases, and the inability of teachers to identify these cases due to their erroneous representations might lead to inducing confusions for students.

Three teachers were able to identify stochastic situations and did not present any misconception. The remaining thirteen teachers failed to identify stochastic independency situations (4.1). A few had recognized that there are cases where they cannot apply the “common sense paraphrase of the definition” after being “trapped”. This is definitely an indicator of an erroneous representation of independent events and of the presence of confusions and misconceptions. Misconceptions were evidenced among ten teachers, and we cannot draw any conclusions for the other three teachers who kept the cautious, careful attitude.

Misconceptions M1 and M2 were assumed to be present respectively with two teachers, who were careful in their work and we couldn't draw any conclusion. As for the others, we were able to classify their misconceptions as confusions:

The first confusion, the explicit confusion between “incompatible events and independent events”, was present among two teachers in 4.2.

The second confusion, the implicit confusion between incompatible events and independent events, was present among six teachers as described in 5.

At this point, we conclude that the confusion between “incompatible events” and “independent events” is present among eight teachers in an explicit or in an implicit form. Might that be a factor inducing the same confusion that appears explicitly among students?

Appendix 1

Problem 4

<p>On lance un dé parfaitement équilibré et on désigne par A, l'événement 'sortie d'un nombre pair', par B l'événement 'sortie d'un nombre supérieur ou égal à 4', et par C l'événement 'sortie d'un multiple de 3'. 1) A et B sont-ils indépendants ? 2) A et C sont-ils indépendants ? <i>A fair die is tossed once. Consider the following events, A: "the number is even", B: "the number is greater or equal to 4", and C: "the number is a multiple of 3".</i> 1) Are A and B independent? Justify your answer. 2) Are A and C independent? Justify your answer.</p>	
<p>S4E1 1) Non, car on a des éléments en commun dans A et B. Même, si dans A on n'a que des nombres pairs, on aura dans B des nombres pairs aussi. 2) On peut avoir un multiple de 3 qui soit un nombre pair tel que 6. Il y a une relation entre A et C ; ils ne sont pas indépendants.</p>	<p>S4E2 1) A: {2; 4; 6} et B: {4; 5; 6}. Non, les événements ne sont pas indépendants puisqu'ils ont 4 et 6 en commun. 2) A: {2; 4; 6} et C: {3; 6}. Non, puisqu'ils ont 6 en commun.</p>
<p>S4E3 1) A et B sont indépendants car si A est réalisé ou non cela n'a aucune influence sur B. Si A réalisé ou non, B peut l'être ou ne pas l'être. 2) Si A est réalisé, C peut être réalisé ou ne pas l'être. Si A n'est pas réalisé, C peut l'être ou ne pas l'être.</p>	<p>S4E4 1) $P(A) = 1/2$; $P(B) = 1/2$; $P(A \cap B) = 1/3$. $P(A) \times P(B) = 1/4 \neq P(A \cap B) \Rightarrow A$ et B ne sont pas indépendants. 2) $P(A) = 1/2$; $P(C) = 1/3$; $P(A \cap C) = 1/6$. $P(A) \times P(C) = P(A \cap C) = 1/6 \Rightarrow A$ et C sont indépendants.</p>

Problem 6

<p>On tire au hasard une carte d'un jeu de 32 cartes. On désigne par E, l'événement : 'tirer un pique', et par F, l'événement : 'tirer une dame'. Les deux événements E et F sont-ils indépendants ? <i>A card is drawn from a deck of cards. E is the event: E: "the card is a Queen" and F, the event: "the card is spade".</i> Are E and F independent? Justify your answer.</p>	
<p>S6E1 Non, car il existe une dame de pique donc il y a une relation entre E et F. D'où, les événements ne sont pas indépendants.</p>	<p>S6E2 Les deux événements sont indépendants car chaque événement désigne une chose différente de l'autre. Si E est réalisé, F peut-être réalisé ou pas. Pas d'influence de E sur F.</p>
<p>S6E3 Si on tire un pique, on peut avoir la dame de pique et on peut ne pas l'avoir ; si on ne tire pas un pique, on peut aussi avoir une dame ou ne pas l'avoir. Donc, on peut avoir E sans F ou F sans E ; par suite E et F sont indépendants.</p>	<p>S6E4 $P(E \cap F) = P(E) \times P(F)$. $P(E \cap F) = \frac{1}{32}$; $P(E) \times P(F) = \frac{8}{32} \times \frac{4}{32} = \frac{1}{32}$. $P(E \cap F) = P(E) \times P(F)$. Donc E et F sont indépendants.</p>

Appendix 2

Question 1 : Comment définissez-vous 2 événements indépendants ? (*How do you define 2 independent events ?*)

Question 2 : J'ai remarqué que parfois, certains utilisent l'expression « événements non indépendants » et d'autres parlent d' « événements dépendants ». Qu'en pensez vous ? Est-ce qu'on peut utiliser l'une ou l'autre des expressions dans l'enseignement ? Pourquoi ?

(*Some teachers use "dependent events" and some others "non independent events". Can we use any of these two expressions while teaching? Why?*)

References

- DUPUIS, C., ROUSSET-BERT, S. : 1998, 'De l'influence des représentations disponibles sur la résolution de problèmes élémentaires de probabilité et sur l'acquisition du concept d'indépendance', *Annales de didactique et de Sciences cognitives*, 6, 67-87.
- MAURY, S. : 1984, 'Tirages probabilistes : une aide pour « donner un sens » aux notions d'indépendance stochastique et de probabilité conditionnelle ?', *Publication de l'IREM de Montpellier*, ISBN 2-909916-42-1
- MAURY, S.: 1985, 'Influence of the question in a task on the concept of independence', *Educational Studies in Mathematics*, 16, 283 – 301.
- MAURY, S. : 1986, 'Contribution à l'étude didactique de quelques notions de probabilité et de combinatoire à travers la résolution de problèmes', *Thèse de doctorat d'état – Montpellier*.
- SANCHEZ, E. : 2000, 'Investigaciones didácticas sobre el concepto de eventos independientes en probabilidad', *Recherches en Didactique des mathématiques*, 20.3, 305 – 330.
- STEINBRING, H. : 1986, 'L'indépendance stochastique', *Recherches en Didactique des mathématiques*, 7.3, 5 – 50.

THE NATURE OF THE QUANTITIES IN CONDITIONAL PROBABILITY PROBLEMS. ITS INFLUENCE ON PROBLEM SOLVING BEHAVIOUR

M. Pedro Huerta, *Universitat de València, Spain*
M^a Ángeles Lonjedo, *IES Montserrat, Spain*

Abstract: *In order to solve verbal conditional probability problems, students are involved in a process in which we can identify several steps or phases. One of them is that of the translation from the text of the problem, generally written in everyday language, to that of mathematics. In translating sentences, students should recognize events and probabilities. But, in much of those problems data are not explicitly mentioned in terms of probability. In this case, students can solve these problems with the help of arithmetical thinking and not necessarily with the help of probabilistic reasoning. Other works (Ojeda 1996, Huerta-Lonjedo 2003) already referred to that but not adequately. In this piece of work we investigate, through an exploratory study of 166 students from different school levels, the extent to which the nature of quantities in conditional probability influences the way in which students solve these problems.*

Introduction

We use the term problem in a Puig (1996) sense, that is, any problematic situation in a school context. Probability problems are problems in which the question is about the probability of an event and conditional probability problems involve at least one conditional probability, either as data and question or both. In this report we consider conditional probability problems written in everyday language.

In addition to the nature of data, there are, as we know, some others factors that also have an influence on problem solving of conditional probability tasks. One of these factors is not necessarily previous knowledge about relationships between probabilities but, for example, the identification of the events and their probabilities. The prior identification of events and the corresponding assignment of their probabilities have to do with semiotic and semantic aspects as well as the right correlation between data and events. In this piece of work we are not going to deal with this issue, instead we investigate the nature of quantities in problems, and its influence on the problem solving process.

When data in conditional probability problems are expressed in terms of frequencies, percentages or rates, students do not necessarily interpret them as probabilities. Consequently, relationships between probabilities are not used when students are solving problems, at least in an explicit way. However, this does not mean that no student can solve these problems. Of course, there are students that succeed in solving, but they mainly use arithmetic thinking and not probabilistic thinking. It is only at the end of the problem solving process that students answer the question in

terms of a required probability, usually by assignment methods. In this paper, we will show the results of an exploratory study with 166 students from different school levels solving conditional probability problems, in which the type of data is varied systematically so that its impact on solving strategies of individuals may be studied.

Nature of quantities in conditional probability problems

From an investigation of conditional probability problems in textbooks we noticed that data involved are not always expressed in terms of probabilities. Previous studies (Ojeda 1996; Huerta, Lonjedo 2003; Lonjedo, 2003) show that problems could be solved just by using numeric thinking. We, too often observed students using arithmetic instead of probabilistic thinking when solving conditional probability problems. This is because data are not being interpreted consciously as probabilities and consequently, students do not need to use relationships between probabilities to solve the problem. It is only at the end of the problem solving process that students try to express their answer in terms of the required probability and assign a probability to their arithmetic solution. We will term this strategy as “solving by numeric assignment”. On the other hand, we will use “solving by probability calculations” to denote the strategy to use probability relations to derive at a solution (see next example)

According to these different strategies, we will subsequently classify probability problems into assignment and calculation problems. For a conditional probability problem this means that it will be classified as assignment problem if the quantities involved are presented as frequencies or percentages, and it will be classified as calculation problem if data involved in the problem are given as probabilities and, consequently relationships between probabilities are needed in order to answer the posed question.

Nevertheless, teachers and textbooks usually present problem situations asking for this type of probabilistic thinking, which is by no means really requested by the posed task. The following example illustrates matters.

Two machines A and B produce respectively 100 and 200 pieces. It is known that machine A produces 5% faulty pieces and machine B produces 6% faulty pieces. If you took a piece at random calculate: a) The probability that this piece would be faulty. b) Knowing that the piece is faulty, the probability that it is made by machine A.

Dos máquinas A y B han producido respectivamente, 100 y 200 piezas. Se sabe que A produce un 5% de piezas defectuosas y B un 6%. Se toma una pieza y se pide: a) Probabilidad de que sea defectuosa. b) Sabiendo que es defectuosa, probabilidad de que proceda de la primera máquina (Cuadras 1983, p. 55).

Solución

Indiquemos por: $M_A = \{ \text{la pieza procede de la máquina A} \}$
 $M_B = \{ \text{la pieza procede de la máquina B} \}$

Entonces,

$$\Omega = \{ 300 \text{ piezas} \} = M_A + M_B$$

$$P(M_A) = \frac{1}{3} \quad P(M_B) = \frac{2}{3}$$

1) Sea $D = \{ \text{la pieza es defectuosa} \}$

$$P(D) = P(D/M_A) \cdot P(M_A) + P(D/M_B) \cdot P(M_B) = (0.05) \cdot \frac{1}{3} + (0.06) \cdot \frac{2}{3} = 0.0567$$

2) Es la probabilidad de M_A , condicionada a la presencia de D

$$P(M_A/D) = \frac{P(D/M_A) \cdot P(M_A)}{P(D/M_A) \cdot P(M_A) + P(D/M_B) \cdot P(M_B)} = \frac{(0.05) \cdot \frac{1}{3}}{0.0567} = 0.2941$$

Fig. 1. Solution of the schoolbook

The solution of this problem can be seen in Figure 1. The textbook considers this as a problem of calculation, consistent with its placement in the unit of “Total Probability and Bayes’ Theorem”. However, some students (Lonjedo, 2003) solved problems similar to that both in data nature and in data structure by the numeric assignment strategy – according to the nature of presented data:

If we have 100 pieces from machine A and 5% are faulty, in 100 pieces we have 5 faulty ones. If we have 200 pieces from machine B and 6% are faulty, in 200 pieces we have 12 faulty ones. In total, among 300 pieces we have 17 (5+12) faulty pieces. So, if we take a piece at random, the probability that it will be faulty is 17 out of 300 or 17/300 or 0.056. We have 17 faulty pieces, 5 of them are from machine A and 12 from machine B. If we know that the piece is faulty, the probability it is made by machine A is 5 out of 17 or 5/17 or 0.2941.

Si tenemos 100 piezas de la máquina A y el 5% son defectuosas: tenemos 5 piezas defectuosas de las 100 de A. Si tenemos 200 piezas de la máquina B y el 6% son defectuosas: tenemos 12 piezas defectuosas de las 200 de B. En total, de 300 piezas de las dos máquinas, 5+12=17 son defectuosas. Luego la probabilidad de ser defectuosa es: 17 de 300 o 0.056. Para la segunda cuestión, tenemos 17 piezas defectuosas, de donde 5 vienen de la máquina primera, luego la probabilidad pedida es de: 5 de 17 o 0.2941.

Different nature of data

Conditional probability problems may be classified according to the nature of the data presented in the formulation of the problem. We distinguish the following types:

Data expressed in probability terms

If quantities are expressed in terms of probability, they quantify the probability of a certain event A by a number $p(A) \in [0,1]$, as in the following example:

Complete the next contingency table. From this table build a tree diagram and calculate $p(B|A)$, $p(\text{noB}|A)$, $p(B|\text{noA})$ and $p(\text{noB}|\text{noA})$ - $p(\text{noB}|A)$ means $p(\bar{B}|A)$.

	A	noA	Total
B	0.4	0.2	

Working Group 5

Completa la següent taula de contingència. A partir de la taula, confecciona un diagrama d'arbre i determina $P(B/A)$, $P(\text{no}B/A)$, $P(B/\text{no}A)$ i $P(\text{no}B/\text{no}A)$ (Matemáticas 4t ESO, p. 240).

noB	0.25		
Total			1

Here, the solution is derived by relationships between probabilities, namely by: $p(A|B)=p(A\cap B)/p(B)$, that is, only by using probabilistic calculations and thinking.

Data expressed in absolute frequency terms

When in a conditional probability problem data are expressed in terms of absolute frequencies, they express the frequency of the objects that satisfy certain characteristics. From a mathematical point of view, frequency can be seen as a cardinal number associated to the set that represents these objects. Consequently, the quantity in a problem presented as frequencies has to be used with that meaning. Thus, $p(A|B)$ is obtained by comparing two numbers: $p(A|B) = n(A\cap B)/n(B)$.

On the other hand, because $A|B$ is not an event, we cannot consider a set that represents it. So, data referring to a conditional probability could never be expressed in terms of absolute frequencies. If we do so, the only meaning that one can associate to such data will be that of a cardinal number associated to an intersection event. The following example illustrates matters:

An intelligence test was administrated to a group of 500 students to assess their academic performance. The results of that test were as follows (contingency table). Let A be “having higher intelligence test” and B “having a higher academic performance” Answer the questions: a) Are A and B independent events? b) If we randomly choose a student with higher performance at school, what is the probability that he/she has higher intelligence?

	Rendimiento académico	
Inteligencia	Alto	Bajo
Superior	200	80
Inferior	100	120

En un grupo de 500 individuos se pasó un test de inteligencia y se midió su rendimiento académico. Los resultados fueron como sigue (tabla de contingencia). Considerando que A es “ser superior en inteligencia” y B es “tener rendimiento alto”, averiguar: a) Si A y B son independientes. b) Si se selecciona al azar un alumno con rendimiento alto, ¿cuál es la probabilidad de que sea superior en inteligencia? (Santos Serrano, 1988, p. 248)

Data expressed in terms of rates

When quantities are shown in terms of rates, data are implicitly expressed in terms of probability and it is up to the solver to decide whether to translate the rates to probabilities or not. Here, we have two examples. The first one shows two rates as data and the second one shows data in percentages:

In Sikinie one man out of 12 and one woman out of 288 are affected with Daltonism. The frequencies of men and women are the same. A person is chosen at random and it is known that he/she is affected with Daltonism. What is the probability that this person is a man?

Working Group 5

En Sikinie, un homme sur 12 et une femme sur 288 sont daltoniens. Les fréquences des deux sexes sont égales. On choisit une personne au hasard et on découvre qu'elle est daltonienne. Quelle est la probabilité pour que ce soit un homme? (Engel 1975, p. 270)

Experience shows that during the process of making circuits for radio transistors 5% of them faulty. A device that is used to find out which are faulty, detects 90% of the faulty ones, but also qualifies 2% of faulty circuits as correct. What is the probability that a circuit is correct if the device says that it is faulty? What is the probability that a faulty circuit is qualified as correct?

En el proceso de fabricación de circuitos impresos para radio transistores se obtiene, según demuestra la experiencia de cierto fabricante, un 5% de circuitos defectuosos. Un dispositivo para comprobar los defectuosos detecta el 90% de ellos, pero también califica como defectuosos al 2% de los correctos. ¿Cuál es la probabilidad de que sea correcto un circuito al que el dispositivo califica como defectuoso? ¿Cuál es la probabilidad de que sea defectuoso un circuito calificado de correcto? (Grupo Cero 1982, p. 170)

Data expressed in combined terms

There are certain conditional probability problems in textbooks where not all data are expressed by the same type, like in the examples above, but by combining more than one. We find data presented both in terms of probability and percentages, probability and rates, or rates and frequencies. In the following problem, for example, data as percentages combines more than one sense.

In a high school class, the percentage of students that succeeded in History (A) was 60%. In Mathematics (B) it was 55%. Knowing $p(A|B)=70\%$, what is the probability that a student chosen at random did not succeed in either topic?

En un curso el porcentaje de aprobados en Historia (A) es 60 %. Para Matemáticas (B) es del 55 %. Sabiendo que $p(B/A) = 70 \%$, ¿cuál es la probabilidad de que, escogido al azar un alumno, resulte no haber aprobado ninguna de las dos asignaturas? (Santos Serrano, 1988, p. 248)

The examples show how data in conditional probability problems are not always expressed in probability terms or with the same type. When this occurs, the solver should have the ability to interpret them according to or different from the desired meaning in the problem. Depending on how the solver actually interprets the data the solving process may imply either numerical or probabilistic thinking.

The empirical study

One of the objectives of our study was to explore how students solved conditional probability problems when data in problems satisfied specific criteria with respect to their structure and nature. Mainly, we were interested in exploring what kind of thinking –arithmetical or probabilistic– students used in solving the problems, in relation to the structure and nature of the presented data.

The test

We prepared a collection of sixteen conditional probability problems with similar data structure, varying the nature of the presented data and the context. All problems had three pieces of data explicitly mentioned in their formulation. For each constellation of context, we designed a pair of problems, one contained the quantities

in terms of percentages, the sibling in probability terms. For example, in problem 1, one can interpret data as follows

$$p(A \cap B) = 30\%, p(A \cap \bar{B}) = 30\% \text{ and } p(B | \bar{A}) = 40\%$$

In its isomorphic problem, no. 9, data are explicitly mentioned as probabilities

$$p(A \cap B) = 0.3, p(A \cap \bar{B}) = 0.3, p(B | \bar{A}) = 0.4.$$

Both problems, however, share the same question —calculate $p(B)$ — and the same context. The collection of problems contained problems from 1 to 8 with data in terms of percentages and problems 9 to 16, with data explicitly expressed in terms of probability: 1-9, 2-10, 3-11, 4-12, 5-13, 6-14, 7-15, and 8-16; see the appendix. Only a Spanish version is given as we think that a translation cannot fully preserve meaning, and semantic and semiotic factors are influential for the problem perception.

Considering time limitations, we asked each student to solve a total of four problems - two from the "frequency type problems and two with probability format of the presented quantities.

The students

The test was administered during student's regular class time. The sample of students that took part in the study was a total of 166 students distributed as follows as follows over school levels and ages:

School	U-FM	HS2-TS	HS2-SS	HS1-TS	HS1-SS	CS4
Age	20	17-18		16-17		15-16
#	10	38	16	38	37	27

U-FM: University students at the Math's College, studying "Didactics of Mathematics"

HS2-TS, SS: 2nd year high school students specializing in TS – technical subjects, or SS – social sciences, which means different competence at mathematics

HS1-TS, SS: 1st year high school students

CS4: 4th year compulsory students

Analysis of results

The results in the tables below are organized according to the following variables: nature of data, number of students who attempted to solve each problem; number of students who succeeded in solving each problem including its distribution over the various school levels; the number of students that did not answer the specific problem; and depending on the reasoning used in problem solving, the distribution of the number of students that succeeded with probability assignment or probability calculation strategies.

Tables 1 and 2 display the number of students who succeeded in solving and also the number of students who did not attempt to solve the problems or similar - expressed as blanks. Information about students who did not complete successfully the

problems is not reported here. In the process of solving we could observe mistakes and misunderstandings of different nature.

	Counts	Frequencies type problems								Probabilistic type problems							
	Problem	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
	Sample	34	33	33	34	66	33	67	34	34	33	33	34	33	33	33	33
	Succeeded	4	6	1	8	20	2	4	0	0	2	0	6	3	2	2	0
Level	UFM	1	0	0	2	1	2	4	0	0	1	0	2	1	1	2	0
	HS2-TS	0	3	0	2	6	0	0	0	0	1	0	1	2	0	0	0
	HS2-SS	0	0	0	3	2	0	0	0	0	0	0	1	0	0	0	0
	HS1-TS	1	1	0	0	8	0	0	0	0	0	0	0	0	1	0	0
	HS2-SS	2	2	1	1	3	0	0	0	0	0	0	1	0	0	0	0
	CS4	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	Blank	5	11	4	2	11	4	12	9	12	4	13	12	11	6	18	19
Strategy	Assignment	4	6	1	7	20	0	1	0	0	0	0	1	0	1	0	0
	Calculation	0	0	0	1*	0	2	3	0	0	2	0	5	3	1*	2	0

Table 1: Number of students succeeded in solving by school level and type of reasoning used – * UFM

	%	Frequencies type problems								Probabilistic type problems							
	Problem	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
	Succeeded	12	18	3	24	30	6	6	0	6	0	18	9	6	6	0	0
	Blank	15	33	12	6	16.6	12	18	27	12	40	35	33	18	55	58	19
Strategy	Assignment	100	100	100	87	100	0	25	0	0	0	16.7	0	50	0	0	0
	Calculation	0	0	0	13*	0	100	75	0	100	0	83.3	100	50*	100	0	0

Table 2 Percentages of students succeeded, blank and, type of thinking of successful students – * UFM

Tables 1 and 2 display the same information, with Table 1 in absolute frequencies while Table 2 gives the results in percentages. It should be noted that for problems 9 to 16, where data are expressed in terms of probability, the percentages of correct solutions is lower than for the first 8 ones. On the other hand, we would like to point out that when data are shown in terms of probability, the number of students trying to solve the problem is much smaller¹ compared to frequency type problems; with the

¹ According to the official curriculum of students for 1st and 2nd year at high school studying social science-humanities option, some conditional probability knowledge is provided. However, this does not always happen. We cannot assure

exception of P2 where the high percentage of “blanks” stands out. This confirms that nature of data is an influential factor in the problem’s solution processes. The summary table below gathers those columns of Table 2 that give evidence of the solution process of the problems, the nature of their data and those that have been successfully completed:

		Frequencies type problems								Probabilistic type problems							
%		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Strategy	Assignment	100	100	100	87	100	0	25	0	0	0	16.7	0	50	0	0	0
	Calculation	0	0	0	13*	0	100	75	0	100	0	83.3	100	50*	100	0	0

Table 3: Percentages of the types of solutions for those problems successfully completed – * UFM

One of the features to be seen from the Table 3 is that in some cases it contradicts the assertion that data expressed in percentages favours the solution of the problem by assignment. We can see, for example, that in problems 6 and 7; the majority of those students that successful solved the problems used probability calculations. However, all the students in this sample belong to the Math’s College, who have more education in the theory of probability.

We can also notice that in some cases, when data are expressed in terms of probability, some students translate those terms into percentages and solve the problem by using arithmetic thinking and assigning probability at the end of the solution process. This is the case in problem 12. One student from the HS1-SS group translated data expressed in terms of probability into percentages. When solving problem 14, another student, from the HS1-TS group, used the same process of translation.

The eight problems with data presented in percentages and the number of students that succeeded in solving them can be seen from Table 4: The results of problem 8 are coherent with the enunciation of the problem and the competence level of the students. The two students from the Math’s College, theoretically provided with a good knowledge of the theory of probability, tried to solve it by using wrong formulas. The high percentage of answers left “blank” is basically due to the way these data are presented: $p(B|A)=70\%$.

that students belonging to these courses were knowledgeable in the subject. Consequently, problems with data expressed in terms of probability are not attempted to be solved very frequently.

%	Frequencies type problems							
	P1	P2	P3	P4	P5	P6	P7	P8
Problem								
Succeeded	12	18	3	24	30	6	6	0
Blank	15	33	12	6	16.6	12	18	27

Table 4: Percentages of problems that give evidence of the success or failure of the solution.

With reference to problem 2, also with a very high percentage left “blank”– 33% – correspond to 8 blank answers and 3 with an unfinished solution. We do not understand the reason why this happens because the enunciation is similar to problems 1 and 3, where the percentages with a blank response are not so relevant. Moreover, if we observe the results of its isomorphic problem no 9, there are only 12% of left blanks.

Conclusions

We suspect that apart from the nature of the data, there are some other factors that have a direct effect on the way students approach conditional probability problems. These factors are not necessarily related to knowledge of relationships between probabilities. The nature of probability problems in textbooks is influential in the classification of the problems: problems of probability assignment or problems of probability calculation. The problems successfully solved by the students who took part in this research could be classified as problems of probability assignment. Most of these students did not understand data as probabilities and consequently would never use the relations between probabilities to calculate the probabilities requested in the problems. Data in these problems are presented in percentages and students solve them by using numerical thinking and final assignment of a probability. However, a few students attempted to solve problems with data presented in terms of probability. In this case, the students mostly approached the problems by using probability calculation rules.

Numerical data shown in selected probability problems, acquire some meaning to the students when they are expressed in terms of percentages. When quantities have specific meaning for students, they can produce new quantities that can also be relevant for the solution of the problem, thus facilitating the problem solving process. However, when quantities are expressed in terms of probabilities they do not make feasible that production of new quantities, mainly if one is not competent with relationships between probabilities or with formulas. Consequently, we will not be saying anything new (Ojeda, 1996) if we continue believing in the solution of probability problems where data suggest a focus on probability as a frequency before this is shown in a formal way. This applies not only to solving probability problems, but also to conditional probability. If conditional probability problems were focused

in the way we suggested, it would allow their inclusion in arithmetic lessons or the use of rates and proportions for their solution, as a prior step to teaching rules or formulas.

References

- Cuadras C.M, (1983) *Problemas de Probabilidades y Estadística*, vol 1: Probabilidades, (Barcelona: Promociones Publicaciones Universitarias)
- Engel, A.,(1975) *L'enseignement des probabilités et de la statistique*, vol. 1. (France; CEDIC)
- Grupo Cero, (1982), *Matemáticas de Bachillerato. Curso 1.* (Barcelona: Teide)
- Grupo Erema, (2002), *Estadística y probabilidad. Bachillerato. Cuaderno 4.* (Madrid: Bruño).
- Huerta, M. Pedro (2003) Curs de doctorat en Didáctica de la probabilitat. Departament de Didàctica de la Matemàtica. Universitat de València (documento interno).
- Huerta, M. Pedro, Lonjedo, M^a Ángeles (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.
- Lonjedo, M^a Ángeles, (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 Ángeles, Huerta, M. Pedro, (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. (A Coruña: Universidade da Coruña).
- Ojeda, A.M. (1996), Contextos, Representaciones y la idea de Probabilidad Condicional, *Investigaciones en Matemática Educativa*, pp. 291-310. (México: Grupo Editorial Iberoamérica).
- Puig, L., (1996), *Elementos de resolución de problemas.* (Comares: Granada).
- Ramírez, A. y otros, (1996), *Matemáticas 4t ESO, Opción B.* (Valencia: Ecir).
- Santos Serrano, D., (1988), *Matemáticas COU, Opciones C y D.* (Madrid: Santillana).

Appendix:

The test problems to the right are equivalent to the left ones in the same line with respect to context.

Working Group 5

Frequency type problems	Probability type problems
P1: De todos los alumnos del instituto, un 30% practican baloncesto y fútbol y un 30% practican el baloncesto y no practican el fútbol. Sabemos que de los alumnos que no practican baloncesto un 40% hacen fútbol. Calcula la probabilidad de practicar fútbol.	P9: En un instituto, la probabilidad de practicar baloncesto y fútbol es 0'3 y la probabilidad de practicar el baloncesto y no practicar el fútbol es 0'3. Sabemos que la probabilidad de que elegido un alumno de los que no practica baloncesto, éste practique fútbol es 0'4. Calcula la probabilidad de practicar fútbol.
P2: Un 30% de los huéspedes de un hotel practican el tenis y el golf y un 30% practican el tenis y no practican el golf. Además conocemos que de los huéspedes que no practican tenis un 40% practican golf. Calcula la probabilidad de que elegido un huésped al azar no practique ni tenis ni golf	P10: En un hotel, la probabilidad de que elegido un huésped al azar éste practique el tenis y el golf es 0'3 y la probabilidad de que practique el tenis y no practique el golf es 0'3. Además conocemos que la probabilidad de que elegido un huésped de los que no practican tenis éste practique golf es 0'4. Calcula la probabilidad de que elegido un huésped al azar no practique ni tenis ni golf.
P3: En una academia de idiomas un 30% de los alumnos estudian inglés y francés y un 30% estudian inglés y no estudian francés. Además, de los alumnos que no estudian inglés, un 40% estudian francés. Calcula la probabilidad de que estudie inglés elegido un alumno que estudia francés.	P11: En una academia de idiomas, elegido un estudiante al azar la probabilidad de que estudie inglés y francés es 0'3 y de que estudie inglés y no estudie francés es 0'3. Además, elegido un alumno de los que no estudian inglés, la probabilidad de que estudie francés es de 0'4. Calcula la probabilidad de que estudie inglés elegido un alumno que estudia francés.
P4: En una empresa el 55% de los trabajadores son mujeres. De las mujeres, el 20% se dedican a las tareas administrativas, y de todos los trabajadores, el 11'25% son hombres y administrativos. Calcula la probabilidad de ser mujer y no realizar tareas administrativas	P12: De los trabajadores de una empresa, la probabilidad de ser mujer es de 0'55. De las mujeres, la probabilidad de dedicarse a las tareas administrativas es de 0'2, y elegido un trabajador al azar, la probabilidad de ser hombre y administrativo es 0'1125. Calcula la probabilidad de ser mujer y no realizar tareas administrativas.
P5: En una universidad el 55% de los estudiantes son mujeres. De éstas, el 20% estudian carreras de letras, y de todos los estudiantes, el 11'25% son hombres y estudian carreras de letras. Calcula la probabilidad de que elegido un estudiante al azar (hombre o mujer) estudie carrera de letras	P13: En una universidad, elegido un estudiante al azar, la probabilidad de que sea mujer es 0'55. De éstas, la probabilidad de que estudien carreras de letras es de 0'2, y elegido un estudiante al azar, la probabilidad de ser hombre y estudiar carrera de letras es de 0'1125. Calcula la probabilidad de que elegido un estudiante al azar (hombre o mujer) estudie carrera de letras.
P6: En un campamento de verano el 55% de los integrantes son niñas. De las niñas, el 20% realizan actividades acuáticas, y de todos los integrantes, el 11'25% son niños y realizan actividades acuáticas. Calcula la probabilidad de que eligiendo un integrante que realiza actividades acuáticas, éste sea niña.	P14: La probabilidad de que los integrantes de un campamento de verano sean niñas es de 0'55. De las niñas, la probabilidad de realizar actividades acuáticas es de 0'2, y elegido un integrante al azar, la probabilidad de ser niño y realizar actividades acuáticas es de 0'1125. Calcula la probabilidad de que eligiendo un integrante que realiza actividades acuáticas, éste sea niña.
P7: Un 60% de los alumnos de un colegio aprobaron filosofía y un 70% matemáticas. Además, un 80% de los alumnos que aprobaron matemáticas, aprobaron también filosofía. Si Juan aprobó filosofía, ¿qué probabilidad tiene de haber aprobado también matemáticas? (Grupo Erema, 2002, p. 26, adaptado para la prueba)	P15: En un colegio, la probabilidad de aprobar filosofía es de 0'6 y la de aprobar matemáticas es de 0'7. Además, elegido un alumno de los que aprobaron matemáticas, la probabilidad de que aprobara filosofía es de 0'8. Si Juan aprobó filosofía, ¿qué probabilidad tiene de haber aprobado también matemáticas?
P8: En un curso el porcentaje de aprobados en Historia (A) es 60 %. Para Matemáticas (B) es del 55 %. Sabiendo que $p(B/A) = 70 \%$, ¿cuál es la probabilidad de que, escogido al azar un alumno, resulte no haber aprobado ninguna de las dos asignaturas? (Santos Serrano, 1988, p. 248, adaptado para la prueba)	P16: En un curso la probabilidad de aprobar Historia (A) es 0'6 y la de aprobar Matemáticas (B) es 0'5. Sabiendo que $p(B/A) = 0.7$, ¿cuál es la probabilidad de que, escogido al azar un alumno, resulte no haber aprobado ninguna de las dos asignaturas?

EXPLORING INTRODUCTORY STATISTICS STUDENTS' UNDERSTANDING OF VARIATION IN HISTOGRAMS

Carl Lee, Central Michigan University, USA

Maria Meletiou-Mavrotheris, Cyprus College, Cyprus

Abstract: *Histograms are among the main graphical tools employed in introductory statistics classrooms in the instruction of the topic of variability of distributions. Hence, it might be expected that students have good understanding of histograms. Recent work in statistics education reveals, however, that students have beliefs about the features of histograms that are different from what is intended by instruction. The purpose of the study was to investigate students' ability to reason about variation in histograms. The article describes the insights gained from the study regarding the tendency among students to use "bumpiness", or unevenness, of a distribution displayed by a histogram as a criterion for high variability.*

Introduction

The histogram is among the main graphical representations employed in the statistics classroom for assessing the shape and variability of distributions. Introductory statistics courses have been traditionally using the histogram both as a tool for describing data and as a means to aid students in comprehending fundamental concepts such as the sampling distribution. In addition to being widely used in the statistics classroom, the histogram is a graphical representation of data broadly used in the media to present information. Thus, it might be expected that students are familiar with this type of data representation. Recent work in statistics education reveals, however, that students are likely to have beliefs about the features of histograms that are different from what is expected by statistics instructors (e.g. Lee and Meletiou-Mavrotheris, 2003).

The current study was designed to closely investigate, in a real-classroom setting, college-level introductory statistics students' reasoning about variation in histograms. A set of carefully selected tasks was used in the study in order to examine, and at the same time support, student reasoning. The article describes the insights gained from the study regarding students' tendency to consider the "bumpiness", or unevenness, of a distribution displayed by a histogram as a criterion for high variability (i.e. to concentrate on the vertical axis of the histogram and base judgments solely on differences in the heights of the bars), a belief often observed among students (Chance, Garfield, and delMas, 1999; Meletiou-Mavrotheris and Lee, 2002; delMas and Liu, 2003).

Methodology

Context and Participants: The site for the study was an introductory statistics course in a four-year college in Cyprus. One of the authors was the course instructor. Class met two times a week, for two hours each time. There were thirty-five students in the

class, most of whom majoring in a business related field of study. Only few students had studied mathematics at the pre-calculus level or higher.

Instruments, Data Collection and Analysis Procedures: At the beginning of the semester, students were given a pretest on graph-understanding to provide a baseline for the study. Each of the questions in the pretest was selected from previous studies in statistics education, mainly to provide a point of reference and comparison for our findings. During the course, a set of carefully chosen, often technology-based, tasks related to the construction, interpretation and application of histograms was collected to examine, while at the same time supporting, students' reasoning about variation when solving statistics problems involving histograms. Some of the students were observed and videotaped while working on the tasks. Completed worksheets of each task were collected from the whole class.

A detailed analysis of the ways in which students approached the tasks was conducted. The method of analysis involved inductively deriving the descriptions and explanations of how the students interacted with the histograms and reasoned about variation through histograms. Using related behaviors and comments within each topic, we wrote descriptions of the students' strategies and actions. These descriptions formed the findings of the study. Those findings that relate to the observed tendency to judge the variability of a distribution based on the "bumpiness" of its histogram are described in the next section.

Results

Pre-assessment: Students' tendency to concentrate on the vertical axis when comparing the variation of two histograms and to base judgments on differences in frequencies among the different categories was evident in the pre-assessment. On the "Choosing Distribution with More Variability" task in Figure 1 (adapted from Garfield, delMas, and Chance, 1999), for example, almost half of the students (45 percent) argued that distribution A has more variability. Student explanations for choosing distribution A (see Table 1) suggest that they shared the belief that a 'bumpier' distribution with no 'systematic pattern' has a larger variability (Garfield and delMas, 1990).

Figure 1
"Choosing Distribution with More Variability" Task

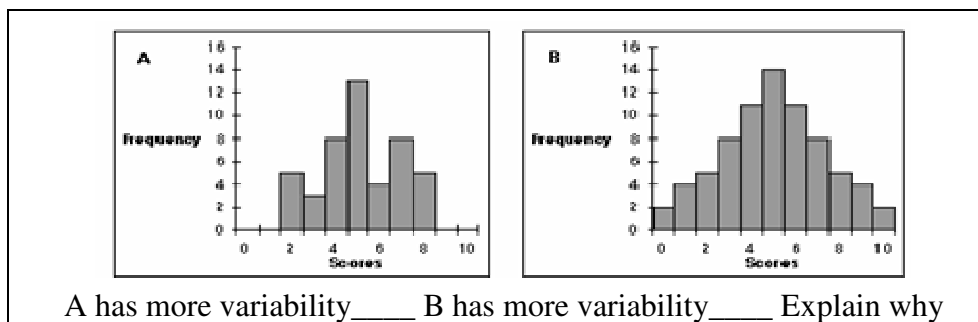


Table 1
Student responses as to “Choosing Distribution with More Variability” Task

Answer	%	Reasons given for response
A has more variability	45% (N=15)	“A has more variability because B is a symmetrical distribution which means all of the values are about the same. In graph A the values are spread, something that makes the variance bigger.” “A is unstable and increases and decreases differently, whereas B increases and decreases at the same rate equally. It is symmetric.”
Graph B has more variability	40% (N=13)	“A has more spread out values.” “A ranges from 1-9, while B ranges from 0-10.”
No response	5% (N=5)	

In a previous study we conducted in a college-level introductory statistics course in the USA (Meletiou-Mavrotheris and Lee, 2002), when giving the same task to students at the beginning of the course, we again found a sizable proportion (26%) arguing that “distribution A has more variability because it’s bumpier.” Chance, Garfield, and delMas (1999), have also found that students often use the “bumpiness” of a histogram as a measure of high “variability”.

Duration of course: The purpose of the study was to closely investigate students’ conceptions of variation in histograms but also to help improve their graph comprehension. In selecting the study activities, we took into consideration findings of a previous study we had conducted, in which we had identified different types of beliefs regarding histograms shared by students which diverge from those intended by statistics instruction (Lee and Meletiou-Mavrotheris, 2003). A set of problems was collected to more closely examine, while also supporting, students’ approaches and strategies when reasoning about variation in histograms.

Next, we provide an example of a typical classroom activity. Our purpose in giving this activity to students was to challenge their tendency to consider the “bumpiness” of a distribution as the main criterion for high variability.

“Value of statistics” Task

In the “Value of Statistics” task (adapted from Rossman, Chance and Locke, 2001), shown in Figure 2, students were given a set of histograms corresponding to the ratings on a 1-9 scale of the value of statistics of students in one of five hypothetical statistics classes, and they had to answer ten questions related to this set of graphs. Students answered the first three questions using traditional means of investigation – paper and pencil – while they answered the remaining seven questions through use of the educational statistical software Fathom[®]. They worked either alone or in pairs to complete the worksheet for the task. Two of the students, Xenia and Ekaterina, were videotaped while working on the task.

Paper and Pencil Stage (Q1-Q3)

All of the students in the class were able to correctly fill out the table in the first question, matching each of the six classes with its corresponding set of frequencies. They did so by contrasting the information in the table with the information provided by the graphical representations.

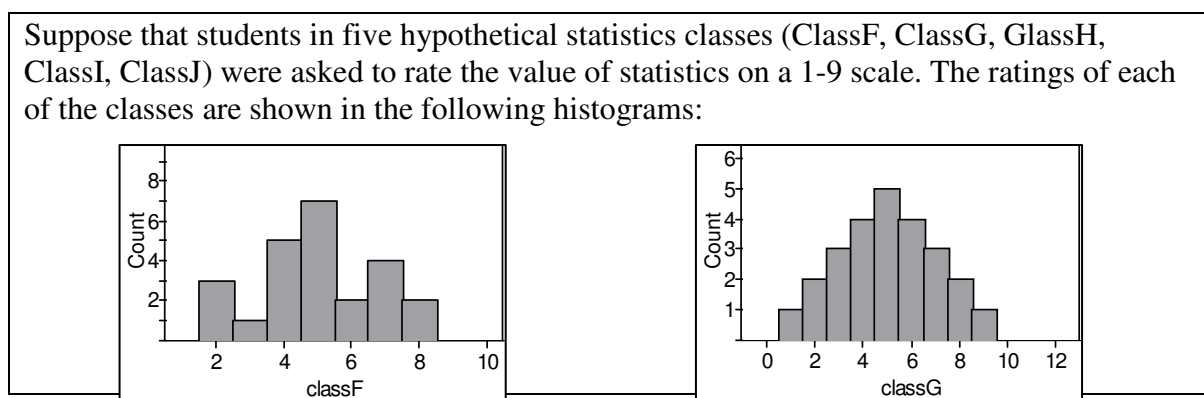
The second question was asking students to use the table and histograms provided to guess which had more variability between classes F and G. This task is similar to the “Choosing Distribution with More Variability” task (Figure 1) students had in the pre-assessment. As we can see in Table 2, the vast majority of students (70%) argued that ClassF has more variability, giving justifications similar to those of students in the pre-assessment who concluded that histogram A had more variability than histogram B. An even higher percentage of students than in the pre-assessment, were equating “bumpiness” of a histogram with high variability.

The tendency to make judgments regarding the variability of a distribution based on the unevenness of its histogram was also evident in students’ responses to Q3, where they had to decide which among classes H, I, and J had the least and which the most variability. Students almost unanimously (93%) argued that ClassJ has the least variability, because “*J is uniform*” (see Table 3). Using a similar mindset, the large majority argued that ClassH has the most variability: “*H has the most variability because the score has a difference of 20 from each other rather than I that has only 10.*”; “*H has the most variability because from 2 it goes to 22 and again to 2, while I from 12 goes 2 and back to 12.*”

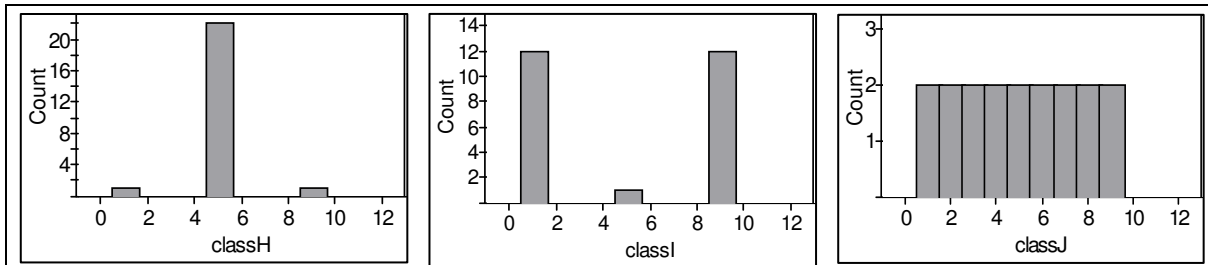
Technology Stage (Q4-Q10)

After completing the first part of the activity, students moved to the computer. They first did Q4. They opened the Fathom[®] file containing the raw data of students’ ratings in each of the classes, drew plots of ratings, and verified that in Q1 they had correctly matched each class with its corresponding set of counts. Next, they used Fathom[®] to calculate the range, interquartile range, and standard deviation of the ratings for each class (Q5). They got the statistics shown in Table 4.

Figure 2
“Value of Statistics” Task



Working Group 5



Q1. The data presented in the histograms is given below in the following table:

Rating	1	2	3	4	5	6	7	8	9
Class ____ count	0	3	1	5	7	2	4	2	0
Class ____ count	1	2	3	4	5	4	3	2	1
Class ____ count	1	0	0	0	22	0	0	0	1
Class ____ count	12	0	0	0	1	0	0	0	12
Class ____ count	2	2	2	2	2	2	2	2	2

Please fill out the table by putting down the name of the class corresponding to each of the five sets of counts in the table.

- Q2. Judging from the tables and histograms, take a guess as to which has more variability between classes F and G.
- Q3. Judging from the tables and histograms, which would you say has the most variability among class H, I, and J? Which would you say has the least variability?
- Q4. Use Fathom and the data in HypoValue.ftm to check that you filled the table correctly.
- Q5. Use Fathom and the data in HypoValue.ftm to calculate the range, interquartile range, and standard deviation of the ratings for each class. Record the results in the table:
- Q6. Judging from these statistics, which measure spread, does class F or G have more variability? Was your expectation in (2) correct?
- Q7. Judging from these statistics, which among classes H, I, and J have the most variability? Was your expectation in (3) correct?
- Q8. Between classes F and G, which has more “bumpiness” or unevenness? Does the class have more or less variability than the others?
- Q9. Among class H, I, and J, which distribution has the most distinct values? Does that class have the most variability of the three?
- Q10. Based on the previous two questions, does either “bumpiness” or “variety” relate directly to the concept of variability? Explain.

Table 2
Student responses to Q2 of the “Value of Statistics” Task

Answer	%	Reasons given for response
ClassF has more variability	70% (N=21)	“F is not symmetrical like G.” “F rises and falls. G increases and decreases at a constant rate.” “F because it has 7 different variables and ClassG only 5 that repeat.”
ClassG has more variability	30% (N=9)	“G covers more values.” “G ranges from 1-9. F doesn’t have as much variation, it scales from 2-8.”

Table 3
Student responses to Q3 of the “Value of Statistics” Task

Answer	%	Reasons given for response
ClassJ has least variability	93% (N=28)	“All the classes show the same frequency.” “J is uniform/rectangular.” “The numbers remain constant. They do not increase or decrease.” “In ClassJ we have the same number of students, it doesn’t matter what score they got.” “The same amount of people gave answers to each category.”
ClassH has least variability	3% (N=1)	(No explanation provided)
H, I, J have the same variability	3% (N=1)	“They have the same variability because they have the same range.”

Table 4
Summary Table for “Value of Statistics” Task

Statistic	ClassF	ClassG	ClassH	ClassI	ClassJ
Range	6	8	8	8	8
IQR	2.5	2	0	8	4
Standard Deviation	1.8	2.04	1.18	4	2.66

Table 5
Student responses to Q6 of the “Value of Statistics” Task

Answer	%	Reasons given for response
ClassF has more variability	13% (N=4)	“ClassF has more variability than ClassG because the interquartile range is more. Our expectation in question B was correct.”
ClassG has more variability	87% (N=26)	“My expectation in B was not correct because Gs values are more spread and now we can see that from standard deviation. As we know, the bigger the standard deviation is, the bigger the spread.” “ClassG has more variability: (i) Range of G is bigger – the bigger the range the more the variability, (ii) Standard deviation is more for G – the bigger the standard deviation the bigger the deviation of values from the mean. Therefore, our answer was wrong.”

In Q6, students had to decide, now based on the statistics they had just calculated, whether their earlier conjecture in Q2 as to which of classes F and G had more variability was correct. As seen in Table 5, the majority of the class, most of whom had earlier argued that ClassF has more variability, now concluded that it is ClassG which has a higher variability. Only four students insisted that ClassF had more variability. Xenia and Ekaterina, our videotaped students, were among those students who had conjectured that ClassF had more variability than ClassG. Looking at the statistics they now decided that their earlier conjecture was incorrect:

Working Group 5

Ekaterina: So the more the standard deviation the more the variability, right? But, by this...judging by the graph we had put down that ClassF has more variability. But by this...by the standard deviation...this says that ClassG has more variability.

Xenia: I don't understand. ClassF has 7 different variables, but ClassG only 5.

Ekaterina: But, here also in H, I, J. J has only one variable but it has more variability than H.

Ekaterina: This graph has numbers on the horizontal axis and count on the vertical axis. So...

Xenia: It is not the count that gives variability. It is the difference between scores. ClassG has more different scores. It goes from 1 to 9, but F goes only from 2 to 8. ClassG has more range. It has more variability.

Ekaterina: So...our expectation was not correct. ClassG has more variability...the range and the standard deviation is higher than the range and standard deviation of ClassF.

Similarly, in Q7 where students had to decide based on the statistics they had calculated, which among class H, I, and J has the most and which the least variability, looking at the statistics made students conclude that their expectations in Q3 were wrong. Whereas, for example, 93 percent of the students had argued in Q3 that ClassJ has the least variability, now 77 percent concluded that it is ClassH which does.

Although in Q6 and Q7 the majority of students based their comparisons of different distributions on typical statistical measures of variability such as the standard deviation and the range, in the last three questions of the task several of them went back into using the “bumpiness” of a distribution as their measure of its variability. In Q8, where they had to decide whether the distribution with the higher “bumpiness” among classes F and G also had higher variability, less than half of the students gave a negative response. Eight students, four of whom had earlier on - after checking the summary statistics - put down that ClassF has a smaller variability than ClassG, ignoring this fact went back into arguing that ClassF has not only more “bumpiness”, but also more variability. Ten students gave no response.

In Q9 students were asked to find which, among classes H, I, and J, has the most distinct values and the most variability. Fifteen students argued that it is ClassH that has the most distinct values, while three others that classes H and I have equally distinct values. These students' interpretation of distinct values suggests that their focus was still on the vertical axis of the histograms.

In the last question, students had to conclude, based on the previous questions, whether either “bumpiness” or “variety” relate directly to the concept of variability. Only fourteen students stated that neither of the two relates directly with the concept of variability (see Table 6). Nine students gave no response. Four students – the same

Table 6
Student responses to Q10 of the “Value of Statistics” Task

Answer	%	Reasons given for response
No	47% (N=14)	“F has more bumpiness but not more variability.” “We can see graphs like H with great unevenness among the number of people that have not much variability.” “None of these concepts directly relate to the concept of variability. Variability describes the spread of the data.”
Yes. More bumpiness implies more variability.	13% (N=4)	“More bumpiness or variety means more variability”
Yes. More bumpiness implies least variability.	10% (N=3)	“Yes, variability is highest for least bumpiness”
No response	30% (N=9)	

that had in the previous question argued that ClassF has more variability than ClassG – reiterated that “*more bumpiness or variety means more variability*”. Three other students drew the over-generalization: “*variability is highest for least bumpiness*”.

Students’ tendency to concentrate on the vertical axis and to judge the amount of variability in a graph based on its “bumpiness” persisted throughout the semester. Even in the end-of-course assessment, when given again the task in Figure 1 of having to decide by looking at the histogram of two distributions of scores which one had more variability, five students (15%) still argued that distribution A has a higher variability because “*it is more uneven than B*”.

Discussion

Our main aim in this study was to strengthen student understanding of one of the most commonly used graphical tools, the histogram. Research indicates that histograms are particularly difficult for students to understand conceptually and cause major problems for many of them (Friel, Curcio, and Bright, 2001). Unlike graphical representations such as scatterplots and time-plots which display raw data, histograms are employed to display the distribution of datasets which have few repeated measures, a large spread in the data, and necessitate the use of scaling of both frequency and data values for purposes of data reduction (Friel, Curcio, and Bright, 2001). People have difficulties in distinguishing the two axes in a display of reduced data such as the histogram (Bright and Friel, 1998). Friel and Bright (1996) found that middle-grade students have difficulties with histograms, in part because of the fact that data reduction leads to a “disappearance” of the actual data. Findings from our study suggest that even college-level students might exhibit a tendency to perceive histograms as displays of raw data. The persistence of the study participants to concentrate on the vertical axis of the histogram when making judgments about the variability of a distribution might be because they perceived each bar as representing

an individual value and not a set of values. In that case, “bumpiness” of the distribution would indeed imply more variability.

Histograms – as well as bar graphs, stem-leaf plots and other graphs – are a transformation from raw data into an entirely different form. Such a transformation changes the data representation – a process that Wild and Pfannkuch (1999) have defined as transumeration – and is one of the fundamental frameworks for statistical reasoning, for better understanding of variation, distribution and many other important statistical concepts. Understanding of this transformation is challenging, and statistics instruction needs to find ways to support it.

Our retrospective analysis of students’ responses to the different instructional tasks we employed during the study, led us to the conclusion that a possible explanation for the persistence of a large group of students in using the “bumpiness” of a histogram as their main criterion for high “variability”, might also be that these students attached meanings to variation that differed from those we had assumed when designing the tasks. When referring to the variability of a distribution of data values in the tasks we assigned to students, the notion of variability we had in mind was that of the spread of values around the center of the distribution, which can be measured using statistical summaries like the standard deviation and the Interquartile Range (IQR). This is the statistical idea of the center of the distribution being the signal (true value, model) and the variation being the spread (noise, residual) around that centre. However, it seems that many of the students in our study did not share this view of variability. Rather, they viewed variability as deviation from an expected pattern. For these students, the signal or model of a distribution was not its centre but its expected distribution (shape), and its variability was the deviation from the expected distribution (Bakker, 2004). As a result, they insisted in choosing distributions with high “bumpiness” as the ones having a higher variability (e.g. distribution A in Figure 1, ClassF in Figure 2), while considering “stable” or symmetric distributions (e.g. distribution B in Figure 1, ClassG and ClassJ in Figure 2) as having little or no variability.

Findings of the study indicate that student knowledge of variation is a much more complex system than instruction often assumes. It encompasses a varied set of ideas not adequately addressed in the statistics classroom. Variability is not at all a precise notation and concept. There are different possible meanings in statistics and everyday language of the terms “variability” and “variation”. In the statistics classroom, variability is usually presented as the spread of values around the center of the distribution. However, as Bakker (2004) points out, in some instances it makes sense to view variability as deviation from an expected distribution. Instruction should not simply dismiss as faulty a student’s tendency to use the “bumpiness” of a distribution as a criterion for high variability. Rather, it should help students differentiate between the different notions of variation and use the appropriate ones depending on the context of the situation.

The current study has provided some useful insights into students' conceptions of the notion of variability as compared to how variability in the statistical sense is measured by certain summary statistics and as is shown in histograms. However, there is still a lot to be learned regarding students' conceptions of variation. Further research should be carried out to investigate the ways in which students perceive variation of a distribution displayed through different means of representation (numerical, tabular, graphical). Findings from such research would greatly enrich our understandings of the ways in which students perceive the idea of variation in different settings and would inform our instructional practices.

References

- Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3(2), 64-83.
- Bright, G. W., and Friel, S. N. (1998). Graphical representations: Helping students interpret data. In P. Lajoie (Ed.), *Reflections on statistics: Agendas for learning, teaching, and assessment in K-12* (pp. 63-88). Mahwah, NJ: Erlbaum.
- Chance, B., Garfield, J., and delMas, B (1999, August). *A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning*. Presented at the 52nd Session of the International Statistical Institute, Helsinki, Finland.
- delMas, R. C., and Liu, Y. (2003). Exploring Students' Understanding of Statistical Variation. In C. Lee (Ed), *Reasoning about Variability: A Collection of Current Research Studies* [On CD]. Dordrecht, the Netherlands: Kluwer Academic Publisher.
- 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.
- Garfield, J., and delMas, B. (1990), Exploring the stability of students' conceptions of probability. In J. Garfield (Eds), *Research Papers from the Third International Conference on Teaching Statistics*, University of Otago, Dunedin, New Zealand.
- Garfield, J., delMas, B., and Chance, B.L. (1999), Tools for teaching and assessing statistical inference: Simulation software. Available at http://www.gen.umn.edu/faculty_staff/delmas/stat_tools/stat_tools_software.htm
- Lee, C., and Meletiou-Mavrotheris, M. (2003). *The Missing Link: Unexpected Difficulties with Histograms*. Presented at the Joint Statistical Meeting. San Francisco, California.
- Meletiou-Mavrotheris, M. and Lee, C. (2002), Teaching students the stochastic nature of statistical concepts in an introductory statistics course, *Statistical Education Research Journal*, 1(2), 22-37. [Available at <http://fehps.une.edu.au/serj>]
- Rossmann, A. J., Chance, B. L., and Locke, R. H. (2001) *Workshop statistics: Discovery with data and Fathom*. Emeryville, CA: Key Curriculum Publishing.
- Wild, C.J., and Pfannkuch, M. (1999), Statistical thinking in empirical enquiry. *International Statistical Review*, Vol. 67(3), 223-265.

IMPROVING STOCHASTIC CONTENT KNOWLEDGE OF PRESERVICE PRIMARY TEACHERS

Dr Robert Peard, *Queensland University of Technology, Australia*

Abstract: *Preservice student teachers (primary) at Queensland University of Technology take only one core Foundations Unit in Quantitative Literacy. One week is devoted to basic introductory probability. However, they are able to select an elective mathematics content subject in probability which extends topics in this field. Many of the students enter this subject with little mathematical background and research by the author is being conducted to determine the most effective way of presenting a unit in more advanced probability to such students. This paper describes some of the difficulties encountered and part of the research examining the use of intuitive, frequentist and axiomatic approaches to probability in the solution of unfamiliar problems requiring a decision making process. Implications for the teaching of probability are drawn.*

The use of probability in decision making is a topic that is now included specifically in many secondary school mathematics curriculums including those of Queensland, Australia (Queensland Studies Authority, 2004). The concept of mathematical expectation has a variety of practical applications and is central to the application of probability in many decision making situations. Earlier research by the author (Peard, 1995) demonstrated that relatively sophisticated applications employing the concept of mathematical expectation can be performed by students with relatively little mathematical background. The author has consequently developed a mathematical content elective unit in probability for B.Ed. primary preservice teachers at QUT based on this research. Recent curriculum developments in school mathematics have seen a much greater emphasis on the role of probability in the classroom worldwide (Borovcnik & Peard, 1997). In Australia, “Chance and Data” features as a strand in the *National Statement on Mathematics for Australian Schools*, (Australian Education Council, 1991) and the Queensland curriculum includes both topics in all years from 4 to 12 (Queensland Studies Authority, 2004). However, instruction in probability has been described as “a very difficult task, fraught with ambiguity and illusion” (Garfield & Ahlgren, 1988, p. 57), and many difficulties in its instruction have been reported in the Australian literature (See, for example, Peard, 1996, 2001; Truran, 1997; Watson & Kelly, 2004; Way, 1997).

The Content of the Unit

The unit begins with an informal approach, building on the students’ intuitive understandings, interest in and familiarity with probability without assuming any prerequisite knowledge other than the ability to convert fractions to decimals and percents, fractional equivalence and basic operations. One of the major objectives of the unit is to ensure that the students do not hold any of the misconceptions about probability that are reported as common. These misconceptions, including the

“gamblers’ fallacy”, are not confined to naive subjects and are prevalent among tertiary students (Peard, 1996; Shaughnessy, 1992). Key to the remediation of these misconceptions is the development of the concept of independence and mathematical expectation. The study of basic probability begins with the examination of the mathematics of simple games of chance and skill with the emphasis on the concept of mathematical expectation and its use in the decision making process in such games. Towards the end of the semester, more complex probabilities are introduced and the content extends to the applications of Binomial and Poisson probabilities using Microsoft Excel, including the use of these techniques in simple hypothesis testing.

Different Approaches to Probability

The literature commonly identifies three different approaches to probability; *classical* (symmetrical or axiomatic), *frequentist* (experimental) and *intuitive* (Shaughnessy, 1992, p. 469). Difficulties are associated with each approach. Most introductory courses in probability begin with situations in which the outcomes are equally likely. In doing this there is an assumption of equal likelihood based on *symmetry* (coins, dice etc.) without the formal recognition of the axioms underlying such assumptions. The *frequentist* approach suffers from the clear difficulty that often short term frequencies give vastly differing results from long term, and to make inductive conclusions in probability is fraught with danger. The process may fail completely. For example, if we wish to show that a "six" on the throw of a single die is as likely as any other number, the short term frequency may give contrary results and reinforce the child's misconception that it is harder to throw a six. Furthermore, earlier research by the author (Peard, 2001) has shown that students using frequentist probabilities often make an assumption of equal likelihood when none exists. Nevertheless it is important to include frequentist probabilities in any course as applications occur in situations where there is no symmetry and the axiomatic probabilities might not be available. The use of the Poisson probability is one such section that is included in this unit. In these situations the mean, np , is estimated from the *frequency* of previous occurrences. Applications include insuring against being struck by lightning and calculating the probability that a large lottery prize will be shared by a number of winners.

It is well documented that probability is the one field in which our intuition is often unreliable (See, for example, Borovcnik & Peard, 1997; Fischbein, Nello & Marino, 1991; Peard, 2001; Shaughnessy, 1992). Nevertheless, there are situations in which we need to use some intuition, based on prior experiences, in the estimation of probabilities. In insurance, for example, where the probabilities are clearly not symmetrical and there may be inadequate data from which to draw frequentist estimates, a degree of intuition is often used.

The Use of Simulation

Simulation can be useful in obtaining a frequentist probability in many situations. While activities of this nature feature in the unit, it is emphasized that care must be

taken with this. In some situations a simulated solution is much easier to obtain than a theoretical one. For example, to calculate the probability of having an opening hand in a game of Bridge, or a pair of Aces in a Poker hand, the use theoretical probabilities would be beyond the scope of this unit. However, with modern computers, or even with enough players dealing cards, a simulated solution is a relatively easy matter. However, simulation does not necessarily lead to or improve any understanding of the underlying theory or analysis of the situation. Borovcnik & Peard (1997), cite the simulated solution to “Monty's Dilemma” (p. 376) as an example of where a correct solution is readily and easily obtained through simulation, but without this resulting in any improved understanding of why.

The Equally Likely Misconception

The assumption of assigning equal likelihood to the outcomes of an event is not always justified. This is, in many instances a misconception. We see this in its simplest form when, for example, young children are asked: “In a class there are 12 girls and 16 boys. The teacher puts the name of each child in a hat and draws one out at random. What is the probability of the name being a boy?” Children who answer $\frac{1}{2}$, arguing that it can be either a boy or girl and that these are equally likely, are exhibiting this misconception. More subtle is the situation where we ask a group of people to select a number from 0 to 9 “at random”. There is a tendency to assume that each of the 10 digits are equally likely and that about $\frac{1}{10}$ of the group will choose each. People are often surprised to learn that the numbers 7 and 3 are much more likely than any of the others. Unlike drawing numbers from a hat, people do not select at “random”.

Mathematical Expectation

The concept of the mathematical expectation of the outcome of an event as the product of its probability and the return or consequence, is one that has a variety of practical applications and is the key concept to the application of probability to decision making in the unit. The decision of an airline to overbook flights, for example, involves computing the various probabilities of the numbers overbooked and forming the product of these and the associated cost of each eventuating. These are then compared with the probability and costs of empty seats. Other applications include insurance, warranties, restaurant overbooking, cloud seeding, and a variety of situations in gaming and betting. The computation of mathematical expectation in different situations forms a fundamental component of the unit.

The research of the present study requires students to attempt solutions using the three different approaches and compares and contrasts their responses.

The Research Study

The author is involved in on going research into the effectiveness of this unit in the improvement of instruction in this difficult field (Peard, 2001). The present study undertaken within the unit and described here continues this process by examining

the use of and relationships between intuitive, frequentist and axiomatic approaches in the introductory unit.

Aims of the Research

At the conclusion of the course of instruction, a study was undertaken first to examine the students' ability to use simple axiomatic probability in the solution of unfamiliar problems involving symmetrical situations and requiring a decision based on mathematical expectation. Second, the research sought to compare intuitive estimations of the same probabilities with the theoretical and with frequentist probabilities obtained through experimentation.

Methodology

The group studied consisted of thirty students enrolled in the probability elective unit. At the end of the semester they were presented with two mathematical problems each requiring a decision of optimal strategy. The students were required to answer each question using the following strategies:

1. Using intuition, the information given and knowledge from prior similar experiences.
2. Using the relative frequencies obtained from playing the appropriate game many times and observing the experimental outcomes, and
3. Computing the mathematical expectations of the various outcomes using classical probability theory and/or the Binomial theorem.

The first two of these strategies were done in a two hour workshop towards the end of the semester. Both questions were considered. The third strategy was given as a take home problem to be discussed in the following week's session. Again, both questions were to be attempted.

The Questions

The questions were selected from situations in the common dice game of Yahtzee, a game of chance with an element of skill in selecting strategies. This game was familiar to the students, having been played previously in the unit. However, the problems as presented were new or unfamiliar to them.

One objective in the play of Yahtzee involves forming the maximum score of the total of the five dice rolled simultaneously. After the first throw the player has the option of holding and scoring any of the five, and re-rolling the others. This procedure is then repeated for a third throw if the player so chooses. Question 1 is related to a decision involved in this.

Question 1

The option "Chance" scores the sum of the numbers showing on the five dice. What is the optimal game strategy to maximize this score. That is, what numbers (1, 2, 3, 4,

5, 6) should be held after each throw in order to maximize the expected score? What is the expected score using this strategy?

Another objective in the play of Yahtzee involves a decision of which numbers to hold and which to throw when there are several options. Question 2 required a decision in one such hypothetical situation.

Question 2

Suppose that a player has rolled on the first throw 1, 3, 2, 2, 6. There are now several options to consider. Assuming that the “Chance” option is not available, consider two other options:

1. Hold the 1, 2 and 3 and throw two dice.

A “4” on either will result in a “short run” of 1, 2, 3, 4 and score 30. A “4 and 5” will constitute a “long run” and score 40. A pair of ones, twos or threes will result in three of a kind scoring 3, 6, and 9 respectively.

2. Hold the pair of “2s” and roll three dice.

Scores will result from: three twos (6), four twos (8), five twos (Yahtzee, 50), or a “full house”, three of any other number (25). A three, four, and five will result in a small straight (30).

Restricting the situation to the second throw only, which of the two options has the greater mathematical expectation?

Strategies

The questions were done sequentially. For each question there were three parts and students were asked to attempt a solution according to the conditions for each part:

Part 1. An Intuitive Strategy. The students were given the rules of the game and asked to formulate an intuitive strategy. They were not to play the game or do any written mathematical computations other than any intuitive mental procedures based on prior knowledge or the recall of prior results of games of chance involving the throwing of dice (Note: As a result of previous experiences in the unit, all students had prior knowledge that the numbers on the roll of a die are equally likely and that the mean or expected score on any one throw of a single die would be 3.5).

Part 2. A Frequentist strategy. For each question the students were then given some time (about 20 minutes) to play the game in groups with discussion, after which they were to decide whether or not they would change their intuitive strategy as a result of the experimental outcomes.

Part 3. Axiomatic Probability. Since the dice are symmetrical and each number is equally likely, it is possible to apply axiomatic probability theory to the problems. Students were next asked if they could calculate the mathematical expectation of each compound event using basic theory. For this they were given one week with the problem to be discussed in the following week’s class.

Class discussion. Following the completion of parts 1 and 2 for each question, an informal class discussion was held in the one workshop in which students had the opportunity to explain and elaborate on their reasoning. Responses to part 3 were discussed in the workshop of the following week.

Responses to Questions and Discussion

Question 1

Intuitive Strategy

From the responses to this, five distinct strategies emerged. Table 1 shows the strategy and the number of students selecting it in order of popularity, with the mathematical expectation computed (by the author) for that strategy.

Table 1

Intuitive strategies to Question 1

Strategy	Number of students	Expected score
1. Hold 4, 5, and 6 on both 2 nd and 3 rd throws	19	23.12
2. Hold 4, 5, 6 on 2 nd ; Hold 3, 4, 5, 6 on 3 rd .	4	22.92
3. Hold 3, 4, 5, and 6 on both	2	21.94
4. Hold 5 and 6 on 2 nd ; Hold 4, 5, and 6 on 3 rd .	2	23.33
5. Hold 6 on 2 nd ; 4, 5, and 6 on 3 rd .	2	21.32
6. No strategy	1	17.5

Discussion. The 19 students selecting Strategy #1 demonstrated some reasonable intuition in that they recognize that the score of “4” was the critical one to decide whether or not to hold. In this case such intuition was probably a result of extensive familiarity with games involving the rolling of dice in the unit and a knowledge that the mean score on a single roll was 3.5. Others unfamiliar with this may not have demonstrated this intuition. Clearly those six students selecting Strategies 2 and 3 demonstrated little intuitive knowledge beyond recognizing that “1”s and “2”s should not be held. However, only two students selected the mathematically correct Strategy 4 indicating that beyond the simple case of a single throw, intuition was of little value in deciding the optimal strategy in the compound situation of two throws. As can be seen from Table 1, the optimal strategy (#4) is to toss any “4” on the first throw, but hold any “4” on the second. The observed results of Table 1 are not unexpected and confirm the results of others in the field that there is no intuitive mechanism for estimating compound probabilities (See, for example, Fischbein et al., 1991; Peard, 1995).

The Frequentist Approach

As a result of playing the game many times and discussing the outcomes, of the 19 students using strategy 1 above, none changed. Of the 4 using strategy 2, three

changed from holding “3” to throwing on the third throw, as did one of the two students using strategy 3. None using strategies 4 or 5 changed.

Discussion. Of the six students who failed to show any intuitive understanding that a “3” should not be held, four changed strategy as a result of the experiment. In these cases the frequentist method helped in improving their strategy. However, in no cases did it lead to a change to the mathematically correct strategy. This is almost certainly due to the fact that the mathematical expectations of the strategies are only slightly different (Table 1) and would require a much greater number of trials, or the use of computer simulation for these differences to become apparent. Furthermore, while the frequentist approach can lead to correct solutions, or, in this case to an improved strategy for some, as Borovcnik and Peard (1997) point out, this does not necessarily lead to or improve any understanding of the situation.

Theoretical Analysis of Question 1

Most of the students attempted a theoretical solution and were able to get to the first stage of computing that the expectation on the first throw is $3.5 \times 5 = 17.5$. Four students were able to continue a path of finding the expected score of each die under different strategies using compound probabilities, and of these, two arrived at a correct solution. This was a more difficult problem than any they would have encountered in the unit and it was not expected that many would be able to complete a full solution.

Discussion. The difficulties in applying basic probability theory to compound situations to compound situations are well documented (Fishbein et al., 1991; Peard, 1995). In this situation, nearly all students were capable of applying the basic theory to the first throw (a simple probability computation), but few were able to proceed with the analysis of the compound situation.

Question 2

Intuitive Strategy

The responses to this with expected scores (computed by the author) for each option, are shown in Table 2.

Table 2

Intuitive Choice for Question 2

Option	Number	Expected Score
1. Hold the 1, 2, 3.	18	11.89
2. Hold the pair of “2”s	11	4.39
3. No difference	1	

Discussion. Although more students intuitively chose the mathematically correct Option 1, 12 others (40%) failed to do so. In this situation the expectations are vastly

different (Table 2) and it might be expected that a greater proportion of these students with some background in probability would be able to make the correct choice intuitively from the information. Many of those choosing Option 2 demonstrated some limited intuition in that they correctly reasoned that the probability of getting at least one “2” on the throw of three die was quite high and considerably greater than getting at least one “4” on the throw of two dice. Furthermore, they reasoned, Option 2 had more ways of scoring something. However, they failed to take into account the much higher scores of Option 1 and to recognise intuitively that the *product* of probability and score, expectation, would be much greater. Again, these results confirm those of Question 1 when it comes to the intuition of compound probabilities.

The Frequentist Approach

As a result of experimentation, it became apparent fairly quickly that Option 1 has the greater expectation. All those who selected this stayed with their choice, while all those that selected Option 2 changed as a result of the experimental results.

Discussion. In this case the frequentist approach clearly enabled the selection of the correct option as the mathematical expectations of the two options were greatly different. However, once again, while the frequentist approach lead to a correct solution, in following class discussions there was no evidence that it lead to the correct theoretical solution.

Theoretical Analysis of Question 2

Of those who attempted a theoretical solution, most were able to compute the expectation for Option 1 using either binomial probabilities or by enumerating all outcomes to conclude that the probability of at least one “4” was 0.31 and that of a “4 and 5” was 0.056 (though two failed to recognize that a 5, 4 gave the same result as 4, 5). From this eight students correctly computed the correct expectation of 11.89, and six of these students completed the more complex computations for Option 2 (expectation 4.39) using binomial probabilities.

Discussion. The use of the binomial theorem proved to be a valuable strategy for many students. Given the mathematical background of most of the students entering the unit, this is evidence of a good deal of success in the unit objectives. Prior experience enabled them to recognize the suitability of the strategy to the situation (It should be noted that these students had not done the algebraic development or proof of the binomial distribution, but had been taught only its application using Excel). Even though only six students were able to perform a complete analysis, it is encouraging that students with relatively little prior mathematical and algebraic background were able to do so.

Conclusions and Implications

Firstly, these two questions would be considered as fairly “advanced” applications of stochastic reasoning. The fact that they can be attempted with some degree of success by a significant proportion of students with limited mathematical background is evidence of a good deal of achievement. However, the results of the study confirm those reported elsewhere in the literature regarding the difficulties in teaching probability. In both questions intuitive ideas proved unreliable. In Question 1 where expectations were not greatly different, frequentist methods did not yield reliable results, whereas in Question 2 where the greatly different expectations were obtained, the frequentist approach did not help understanding of the situation. The results to both questions highlighted that compound probability is an inherently difficult topic for which there is little natural intuition. Nevertheless, it is contended that all three aspects of probability must be included in a comprehensive and practical unit in the subject. To this effect, the implications of the study are that the employment of games of chance and skill such as the dice game Yahtzee and the analysis of its strategies are effective methods of increasing student awareness of these difficulties and of developing the use of mathematical expectation in decision making.

In particular, the results of Question 1 highlight the danger of reliance on intuition even when the subjects have some theoretical background. A theoretical analysis of the situation based on axiomatic probability was necessary to show the fallacy of the intuitive response to “hold 4’s” on the first throw based on knowledge of the mean of 3.5. In this situation a frequentist approach did not produce reliable results.

The results of Question 2 also highlight the need for an axiomatic analysis. Even when the frequentist approach yields reliable results, it does not necessarily lead to an understanding of the situation. It is contended from the results of this research that many students with relatively weak algebraic and overall mathematical abilities are nevertheless able to use appropriate software to perform meaningful analyses of binomial probabilities in simple situations. The implication is that courses in elementary probability may well be able to include Binomial and Poisson probability applications without extensive theoretical backgrounds.

References

- Australian Education Council (1991). *A National Statement on Mathematics for Australian Schools*. Curriculum Corporation. Canberra. Australian Government Printer.
- Borovcnik, M. & Peard, R. (1997). Probability. In J. Kilpatrick (Ed.) *International Handbook of Mathematics Education*. (pp. 371-401). Dordrecht: Kluwer.
- Fischbein, E., Nello, M., & Marino, M. (1991). Factors affecting probabilistic judgements in children and adolescents. *Educational Studies in Mathematics*, 22(6), 523- 549.

- Garfield, J. B., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education*, 19(1), 44-59.
- Peard, R. (2001). Misconceptions in probability held by primary and early childhood preservice teacher education students. In M. A. Clements & H. Tairib (Eds.) *Energising Science, mathematics and technical education for all*. (pp. 227-237). University of Brunei: Brunei Darussalam.
- Peard, R. (1996). Difficulties with teaching probability. *Teaching Mathematics*, 21(1), 20-24.
- Peard, R. (1995). The effect of social background on the development of probabilistic concepts. In Bishop, A. (Ed.) *Regional collaboration in mathematics education*. (pp. 561-570). Melbourne: Monash University.
- Queensland Studies Authority (2003). Years 1 to 12 mathematics syllabus. Brisbane. Queensland Studies Authority.
- 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* (pp. 465-494). New York: MacMillan.
- Truran, K. (1997). Beliefs about teaching stochastics held by primary pre-service teaching students. In In F. Biddulph & K. Carr, (Eds.) *People in mathematics: Mathematics Education Research Group of Australasia 20th Conference Proceedings* (pp. 538-545). Waikato, New Zealand: Mathematics Education Research Group of Australasia.
- Watson, J., & Kelly, B. (2004). A two year study of students' appreciation of variation in chance and data in the curriculum. In I. Putt (Ed.). (pp. 573-580) *Mathematics Education for the third millennium. Proceedings of the 27th Annual Conference of Mathematics Education Research Group of Australasia*. Townsville. James Cook University.
- Way, J. (1997). Which jar gives the better chance? Children's decision making strategies. In In F. Biddulph & K. Carr, (Eds.) *People in mathematics: Mathematics Education Research Group of Australasia 20th Conference Proceedings* (pp. 568-576). Waikato, New Zealand: Mathematics Education Research Group of Australasia.

RANDOMNESS IN TEXTBOOKS: THE INFLUENCE OF DETERMINISTIC THINKING

Pilar Azcárate, *University of Cádiz, Spain*
José M^a Cardeñoso, *University of Granada, Spain*
Ana Serradó, *University of Cádiz, Spain*

Abstract: *This work presents findings about the influence of deterministic thinking in the probability curriculum given in Spanish Secondary School textbooks (ages 12-16). This influence is studied in two different fields: the analysis of the underlying models of intervention and the definitions of randomness proposed in the textbooks. The methodological strategy used is content analysis, with a sample of four Spanish publishing houses, including five books for each publisher. This content analysis is based on a theoretical proposal that clarifies the influence of teachers' epistemological conceptions of their professional development to construct their body of professional knowledge.*

1. Introduction

From the beginnings of the nineties there have been different reform proposals for changing how probability is presented in Spain's mathematics curriculum. The introduction of these new curricula affects the knowledge about probability, and the teaching and learning perspective. These reform proposals are associated with a constructivist point of view –they introduce a new perspective for interpreting how a student learns and a new role for the teacher and student in the teaching and learning process. But teachers do not motivate the learning process blindly as workers. They interpret and apply the official curriculum with their own personal criteria, in which they empathize their conceptions (Carrillo, 2000).

We use the term “conceptions” in reference to a general mental structure that includes “beliefs, concepts, meanings, rules, mental images and preferences, conscious or unconscious” (Thompson, 1992, p. 132). From a constructivist perspective, conceptions are at the same time “tools” to interpret reality and move in it, and “barriers” to the adoption of other perspectives (Porlán, Rivero and Martín del Pozo, 1997). There is a clear relationship between teachers' conceptions and their experiences during the development of the teaching and learning process. In fact, this relationship makes the conceptions' evolution more difficult (Carpenter and others, 1999).

The difficulty of the evolution of teachers' conceptions refers to teachers' experience and to their knowledge of the discipline of probability. In particular, Fischbein and Schnarch (1997) establish that changes in people's conceptions are not easily produced, affirming the robustness in the conceptual field. Cuesta (2004) argues that a change in teachers' conceptions is constrained by a simplified vision of the teaching and learning process and by an absolutist conception of knowledge. Research has

indicated that the relationship between teachers' conceptions of the nature of science and their classroom practice is complex, and it is mediated and constrained by other factors. These include pressure to cover content, classroom management and organizational principles, students' abilities and motivation, institutional constraints and teaching experience (Bell, Lederman and Adb-El-Khalick, 2000). This idea reaffirms Shulmans' key point, confirming that knowing how to teach particular ideas in science effectively is not solely a pedagogical question; it is also impacted by the nature of the subject matter. Firstly, because different ideas in science, and their relationship to other ideas the students may know, present different opportunities for the design of teaching and learning activities. Secondly, because teachers intending to make complex science more accessible to their students might actually tend to distortion and over-simplification (Barnett and Hodson, 2001).

Summing up, the reforms of the Spanish curriculum propose a change in the nature of pedagogical knowledge from absolutist to constructivist perspectives. In order to study the influence of absolutist perspectives on the probability teaching and learning process, it is necessary to introduce a theoretical proposal that will help one understand the complexity of the influences of teacher's conceptions in this process.

2. Teachers' Professional Knowledge about Probability: Hypotheses of Progression

The theoretical proposal introduced is Teachers' Professional Knowledge, defined as *“the ensemble of all the knowledge and experiences that a teacher has and uses in the development of his or her educational work, and that is constructed from the beginning of his or her teacher education until the end of his or her professional career”* (Wamba, 2001, p. 11).

One of the fields comprising Teachers' Professional Knowledge is the metadisciplinary field—i.e., knowledge about the nature of the knowledge. The perspectives considered in this investigation about the nature of science are epistemological and ontological, which have implications for teachers' decisions about what and how they must teach. One of these epistemological and ontological perspectives considered is scientific determinism.

Scientific determinism refers to a doctrine of the world's structure in which any outcome can be rationally predicted, to any desired degree of precision, as long as one has a precise enough description of past outcomes together with all nature's laws (Popper, 1996). This perspective reinforces the necessity of knowing nature's laws to predict the possibility of a determined outcome. In this sense, scientific determinism can be included in an positivist epistemological perspective that encompasses an absolutist vision of truth and knowledge.

From an ontologically deterministic point of view, everything has a determined cause and a determined effect. In order to analyse effects and causes, Descartes proposed an analytical method that consisted in the decomposition of a study object into its parts; this method is termed reductionism. He proposed the reduction of all physical

phenomena to exact mathematical relationships, and suggested the idea of certainty in scientific knowledge. People have rationalized processes in seeking this certainty that conforms to scientific laws and knowledge, developing a hypothetical-deductive or an empirical-inductive method.

A hypothetical-deductive method suggests the existence of an external and closed body of knowledge that is determined by its parts which are hierarchically structured by a cause and effect relationship. Teachers constrained by this conception think there is a unique and real body of knowledge that must be transmitted in a linear, fragmented, hierarchical manner coherent with this hypothetical-deductive method. These conceptions support a **traditional didactic model** of intervention in the teaching and learning process (Porlán and Rivero, 1998). Teachers do not reflect about what and how they must teach, because this is externally determined by mathematical tradition. But, when they begin to reflect about the efficacy of this teaching and learning process, they also argue about this rationalist and deterministic conception of science and the necessity to change their teaching.

The aims of this teaching are to deduce strategies to calculate probability. The conceptual notions introduced to facilitate these calculations are random experiments whose outcomes form a field of knowledge following the Laplace rule (Serradó, 2003). This classical Laplacian tendency is presented to the students in a linear, fragmented, and hierarchic structure based on the teacher's theoretical explanations. The predominance of the teacher's work implies reducing students' tasks to solving individual pen and paper activities of theory application to calculate probabilities in the random games' context.

We consider that teachers may be included in another model, termed the **innovation model**, when they consider the student's active role in the teaching and learning process. Teachers having this tendency do not change their rationalist, absolutist, and determinist conception of science; they only rethink how they can teach. Underlying the consideration of "how to teach" there is an empirical-inductive perspective. This means that the teacher thinks that observation allows the truth to be discovered and objective knowledge to be created through the application of the inductive process (Serradó, 2003).

If the teacher believes that individuals use trial and error routines or orderly systematic approaches to observe, experiment, and collect facts that could support future decisions, then he or she fits a **technological model**. Studies on this model, also denoted the product-process perspective, affirm that the teachers' conceptions affect in turn the students' conceptions of the nature of science, and influence the teachers' behaviour and the atmosphere in the classroom (Mellado, 1997). The aim of teaching is to induce strategies to calculate probabilities. The content introduced consists of random experiments, whose outcomes are used as a tool to calculate probability considered as the ultimate stable tendency of the frequencies (Azcarate, Cardeñoso and Serradó, 2003). The methodological framework proposed is an inductive tendency to construct meanings, with a closed series of theoretical

presentations and activities mainly designed by the teacher to evaluate primary students' intuitions, exploring randomness, the characteristics of probability, and theoretical generalizations. The planned process could allow students to overcome their conceptual errors, fallacies, and paradoxes related to the comprehension of probability.

In contraposition to this tendency, if the teacher believes that knowledge is obtained directly by experience, contact with ideas, and the phenomena of everyday life, then the teacher's tendency fits a **spontaneous model**. School-level knowledge is a flexible open product that arises from the necessities expressed by the students. Teachers included in this tendency think that knowledge is determined directly by observation, which acts externally without intervening with reality. The lack of verification and formalization of knowledge makes its discovery relative to each individual, and not collectively determined.

This is an unsystematic organization of activities to explore the properties of random experiments, outcomes, and probabilities, using all kinds of available resources such as random number generators or computer simulations. This organization is planned in conjunction with students' expectations, but without considering their conceptions. This process can reinforce the use of heuristics and biases when students explain the meaning of probability.

From this perspective, knowledge is relative to each individual of the scientific community, i.e., relative to the observer. Relativism is based on the principle that science is a social and historically conditioned activity. Science is not understood as existing in a closed system, both because of its multiple, tentative, and variable content, and because of its method, which is plural and dynamic (Faerna, 1997). From a constructivist perspective, knowledge can only be constructed by an aptitude for reflection, in which knowledge is itself observed and becomes a fact of knowledge (Morin, 1994).

Morin's words translated to the school ambit mean that teacher and student search for the truth. Both are the subject and object of this search; both must observe and be observed; both criticize and are criticized. Teaching-and-learning is a social process of communication between actors — communication that causes the negotiation of meanings, and the progressive transfer of control and responsibility for the teaching and learning process to the students (Jorba, 1998).

The aim of this model is to construct the meaning of chance, randomness, and the notion of probability. Students cooperate to solve problems related with natural and social random phenomena; while the teacher acts as a mediator helping students to plan and generalize the solution. In this process, students must overcome obstacles to solve the problems, in a process of knowledge reconstruction.

The adoption of this theoretical proposal based on Professional Knowledge allows one to understand the relationships between the nature of science and the different tendencies in the teaching and learning process, from the perspective of teacher

intervention. Professional Knowledge is a system of ideas with different levels of particularization and articulation that is always under reorganization in an open and irreversible process. These different levels imply more information and another concept of the nature of science and pedagogical content that allows a progress hypothesis of Professional Knowledge to be introduced that gives meaning to the evolution of the models of intervention.

This table presents a set of indicators of the categories with which to analyse the content of Spanish textbooks, and to draw conclusions regarding the tendencies in the nature of the underlying knowledge.

Models	Traditional	Technological	Spontaneous	Investigative
Aims	Deduce calculation strategies	Induce strategies to calculate	Explore properties	Construct notions
Contents	Random exp., outcome, Laplace rule	Random exp., outcome, Laplace rule, frequency stability		Phenomenon, chance, randomness, probability
Method/ activities	Linear explanations/ pen and paper	Closed sequences	Unsystematic organization / exploration	Problem solving overcoming obstacles
Context	Random games		Games, computer simulations	Natural, social phenomenon

3. Research Method

Theoretical proposals establish tendencies in which determinism influences the aims, the contents, and the methods developed by teachers. Azcárate, Serradó and Cardeñoso (2004a) conclude that the textbooks provide the knowledge of probability, and the strategies that facilitate the planning and development of teaching. A first approach to the study of the influence of deterministic thinking on the teaching and learning process could be to make a content analysis of the textbooks. The inclusion of this influence in the study of the nature of pedagogical content and of the nature assigned to probability in Spanish Secondary Education textbooks (ages from 12 to 16) only has meaning in an overall study of “the treatment of chance”.

The samples used are the textbooks of four Spanish publishing houses: Bruño, Santillana, McGraw Hill, and Guadiel. The textbooks which comprise the curricular

project of each publishing house are those for the first, second, third, and fourth (option A and B) years.

Previous work about the “the treatment of chance in textbooks” (Serradó, 2003) contributed information about the structure of the units (Serradó and Azcárate, 2003), the description of the probabilistic content (Azcárate, Cardeñoso and Serradó, 2003), and the obstacles to the construction of the knowledge of probability (Azcárate, Serradó and Cardeñoso, 2004). The methodological strategy used in order to clarify this nature is content analysis, referring to two levels — what is manifest or apparent, and what is latent or underlying.

The latent content analysis can inform us about the underlying conceptions of the textbooks' authors, whos used their professional knowledge about probability in writing their textbooks. In order to develop this latent content analysis, the following theoretical proposal is necessary: that the teachers' epistemological conceptions influence their professional development in constructing their professional knowledge.

4. Determinism in Textbooks

In this communication, the results presented are those related to the influence of determinism on the introduction of randomness. Textbooks do not include any section presenting the meaning of randomness: they only include a brief explanation and activities related to random experiments. Focusing the investigation on random experiments, we developed this study at two levels. Firstly, the description of how this notion is included in the textbooks, and secondly, the analysis of the definitions, examples, and activities that are presented.

In all the textbooks, there is a closed section that is found in the other textbooks of the same publishing with no real attempt at reorganization, but simply rephrased with more or less similar words and examples. Also, it is closed because there is no relationship between this section and the others, such as probability. This is not an isolated fact. all the textbooks have a structure in which each concept is developed in a closed, independent section.

All the books of the sample present the deterministic and random experiments in the same section. The Bruño and Guadiel textbooks present first the definition of a deterministic experiment and then the definition of a random experiment. McGraw Hill and Guadiel present these concepts in the opposite order. Perhaps this means that the comprehension of one concept must be complemented by the comprehension of the other. For example, Santillana (4^oA, pp. 264) defines a deterministic experiment as: “*An experiment is deterministic if it is possible to predict the result*”. And, in contraposition, a random experiment is defined as: “*An experiment is random when it is impossible to predict the result*”.

These two definitions are antagonistic. The random experiment is defined as a negation of the deterministic one. Therefore, it makes sense to reflect as to whether the textbook is really presenting a definition of random experiment or if it is actually

defining a non-deterministic one. The conclusion is that textbooks make random and non-deterministic experiments isomorphic.

The possibility of establishing this erroneous isomorphism lies in the use of the words “possible” and “impossible”. The word “possible” means the possibility of determining one of an infinity of results. And the word “impossible” means that no one of these infinite results can be predicted by theory. Both refer to uncertainty in determining this “possibility” and “impossibility” to explain this contingency with deterministic laws. It is in this contingency where Chance is present. Chance reflects the ignorance of the individual as a thinking observer of reality. This ignorance may be related to the impossibility of knowing exactly the initial conditions of the phenomenon under study (Wagensberg, 1998).

The concept of Chance as ignorance can be found in such sentences as: *“In daily life, we find many situations of uncertainty. In a football match between a first division team and another of the third division, the first division team will probably win, but we can not affirm that — we are not sure of it”* (Santillana, 2º, pp. 254).

As in the definition there is no reference to initial conditions or to the presence of Chance, the comprehension of the meaning of these random experiments is reduced to knowing what is not determined. To solve this problem, the Guadiel textbook (4º, pp. 232) defines random experiments by including the concept of chance: *“Random experiments are the ones whose result depends on chance and can not be calculated previously”*. Or the Bruño textbook (3º, pp. 266), which solves the problems introduced by the conditions of the experiment as: *“We denote by random phenomena those phenomena which when occurring under the same conditions can have different results”*.

Only the books of the Bruño publishing house present the definition of random phenomenon. This word emphasises the existence of an external observer who knows both the initial conditions and the results. And he or she applies the laws of nature to determine whether the experiment is deterministic, from a deterministic point of view. The McGraw Hill books complement the definition of random experiment with some examples of random phenomena.

“But it is not necessary to appeal to games to find random phenomena. So, for example, when a couple wants to have a baby they do not know in advance its sex: chance is responsible for joining the chromosomes...” (McGraw Hill, 2º pp. 249).

The textbooks also include a series of exercises to help students understand the difference between these notions. The Bruño and Santillana textbooks include only pen and paper exercises to distinguish between deterministic and random experiments. For example: *“Show which of the following experiments are random: (a) Mixing coffee and sugar, (b) casting two dice and noting down the sum, (c) playing cards, (d) tossing three coins and noting the results”* (Santillana, 2º pp. 254).

The Guadiel and McGraw Hill textbooks also include exercises to observe the properties of random experiments: *“Take a die and cast it 30 times, noting down the sides that appear. Without casting it again, which side do you think is going to appear? Throw the die and observe what happens”*. (McGraw Hill, 1^o pp. 239).

All the deterministic examples and activities presented in the textbooks refer to physical and chemical experiments. In order to determine if these experiments are deterministic or not, the students must know some laws of science. If a student is ignorant of one of these laws, he or she can confuse a random with a deterministic experiment. Most of the examples of random experiments are games of chance obtained with random number generators, such as dice or coins. The students determine that these experiments are random because the causes are unknown and a product of chance. Underlying this impossibility of predicting the causes there resides a deterministic principle.

5. Conclusions

In conclusion, the presentation of randomness in textbooks is influenced by deterministic principles, such as the reduction of its study to the introduction of random experiments. These random experiments are defined by contrasting them to deterministic ones. This definition introduces an erroneous isomorphism between random and non-deterministic experiments. Underlying this definition is a comprehension of uncertainty which gives epistemological meaning to the concept of chance. In order to decide whether an experiment is deterministic or random, the students must apply a law of nature I or conclude that they do not know the causes, both actions being influenced by a deterministic principle.

Previous studies (Serradó and Azcárate, 2003) establish the existence of two tendencies in textbook structure. That of Bruño and Santillana belongs to a traditional tendency with a linear, hierarchical, deductive style of presenting the theoretical content. Some application exercises complement this theoretical content. The students do not need to apply any resources to solve these exercises — they only need their pre-existing ideas about the experiment. The McGraw Hill and Guadiel textbooks present a closed structure of exploration activities, theoretical conclusions, and validation activities. These books correspond to a technological tendency. Their theoretical framework suggests that the epistemological and ontological concepts of the authors of these two textbooks were influenced by deterministic principles. Furthermore, when teachers select one of these books to use in the classroom, they are expressing agreement with this structure. Perhaps they are not reflecting about the influence of determinism, but they are thinking about the coherence of the book with what they think must be taught and how, as determined by mathematics tradition (Azcárate, Serradó and Cardeñoso, 2004a).

If we want teachers' conceptions about randomness to evolve, it will be necessary for them to overcome the constricting use of textbooks as their main source of information. And Spanish textbooks should include a wide variety of examples of

random experiments. Indeed, developing these examples would be an interesting research project.

Appendix

Bruño 3º: Miñano, A. and Ródenas, J.A.: 1998, *Matemáticas 3º*, Editorial Bruño, Madrid, Spain.

Guadiel 4º: Fuster, M. and others: 1996, *Matemáticas, 4 (B)*, Guadiel-Grupo Edebé, Sevilla, Spain.

McGraw Hill 1º: Pancorbo, L and others: 1995, *Matemáticas 1*, McGraw-Hill/Interamericana de España, Madrid, España.

McGraw Hill 1º: Becerra, M^a. V. and others: 1996, *Matemáticas 2*, McGraw-Hill/Interamericana de España, Madrid, España.

Santillana 2º: Almodóvar, and others: 1999, *Matemáticas. Curso 2º ESO*, Grupo Santillana Ediciones, Madrid, España.

Santillana 4º: Almodóvar, J.A., Gil, J. and Nortes, A.: 1998: *Matemáticas Opción A. Curso 4º ESO*, Grupo Santillana Ediciones, Madrid, España.

References

Azcárate, P., Cardeñoso, J.M. and Serradó, A.: 2003, 'Hazard's treatment in Secondary School', *Proceedings of the Third Conference of the European Society for Research in Mathematics Education*, Bellaria, Italy. <http://www.dm.unipi.it/~didattica/CERME3/proceedings>.

Azcárate, P., Serradó, A. and Cardeñoso, J.M.: 2004, 'Obstáculos en el aprendizaje del conocimiento probabilístico: la noción de azar y aleatoriedad', *Proceedings of the XII CEAM*, Huelva, Spain.

Azcárate, P., Serradó, A and Cardeñoso, J.M.: 2004a, 'Las fuentes de información como recurso para la planificación', *Actas del Octavo Simposio de la Sociedad Española de Investigación en Educación Matemática*, La Coruña, Spain, pp. 165-173.

Barnet, J. and Hodson, D.: 2001, 'Pedagogical Context Knowledge: Toward a Fuller Understanding of What Good Science Teachers Know', *Science Teacher Education*, 85, 426-453.

Bell, R.L., Lederman, N.G. and Abd-El-Khalick, F.: 2000, 'Developing and Acting upon One's Conception of Nature of Science: A Follow-Up Study', *Journal of Research in Science Teaching*, 37 (6), 563-581.

Carpenter, T.P., Fennema, E., Franke, M.L., Levi, L. and Empson, S.B.: 1999, *Children's mathematics: Cognitively Guided Instruction*, Heineman, Portsmouth.

Carrillo, J.: 2000, 'La formación del profesorado para el aprendizaje de las matemáticas', *UNO*, 24, 79-91.

- Cuesta, J.: 2004, *La formación del profesorado novel de Secundaria de Ciencias y Matemáticas. Estudio de un caso*, Doctoral Dissertation, UMI's Proquest Digital Dissertations, Michigan.
- Faerna, A.M.: 1997, 'Racionalidad científica y diversidad cultural', *Claves: teoría de la ciencia*, 78, 61-68.
- Fischbein, E. and Schnarch, D.: 1997, 'The evolution with age of Probabilistic, Intuitively Based Misconceptions', *Journal for Research in Mathematics Education*, 28, 96-105.
- Jorba, J.: 1998, 'La comunicació i les habilitats cognitivo lingüístiques', in Jorba, Gómez and Prat (eds.), *Parlar i escriure per aprendre*, ICE/UAB, Barcelona, pp. 130-170.
- Morín, E.: 1994, *Introducción al pensamiento complejo*, Kairós, Barcelona.
- Popper, K.R.: 1996, *El universo abierto. Un argumento a favor del indeterminismo. Post Scriptum a La lógica de la investigación científica. Vol. II.*, Tecnos, Madrid.
- Porlán, R., Rivero, A. and Martín del Pozo, R.: 1997, 'Conocimiento profesional y epistemología de los profesores. I: teoría, métodos e instrumentos', *Enseñanza de las Ciencias*, 15 (2), 155-171.
- Porlán, R. and Rivero, A.: 1998, *El conocimiento de los profesores*, Serie Fundamentos, Colección Investigación y Enseñanza, Diada Editora, S.L., Sevilla.
- Serradó, A.: 2003, *El Tratamiento del Azar en Educación Secundaria Obligatoria*, Doctoral Dissertation, UMI's Proquest Digital Dissertations, Michigan.
- Serradó, A. and Azcárate, P.: 2003, 'Estudio de la estructura de las unidades didácticas en los libros de texto de matemáticas para la educación secundaria obligatoria', *Educación Matemática*, 15(1), 67-98.
- Thompson, A. G.: 1992, 'Teachers' beliefs and conceptions: A synthesis of the research', in D.A. Grouws (ed.), *Handbook of research on Mathematics teaching and learning*, MacMillan, New York, pp. 127-146.
- Wagensberg, J.: 1998, *Ideas sobre la Complejidad del Mundo*, Mathemas 9, Tusquets Editores, S.A, Barcelona.
- Wamba, A.: 2001, *Modelos didácticos personales y obstáculos para el desarrollo personal*, Doctoral Dissertation, Universidad de Huelva, Spain.

Papers on Probabilistic Thinking

Papers by:

Dor Abrahamson

Rolf Biehler

Michele Cerulli

Kjærland Iversen

Efi Papanistodemou

Theodosia Prodromou

PROBLAB GOES TO SCHOOL: DESIGN, TEACHING, AND LEARNING OF PROBABILITY WITH MULTI-AGENT INTERACTIVE COMPUTER MODELS

Dor Abrahamson, *Northwestern University, Evanston, IL, USA*
Uri Wilensky, *Northwestern University, Evanston, IL, USA*

Abstract: *ProbLab, an experimental middle-school unit in probability and statistics, includes a suite of computer-based interactive models authored in NetLogo (Wilensky, 1999). We explain the rationale of two of the models, Stochastic Patchwork and Sample Stalagmite, and their potential as learning supports, e.g., the temporal–spatial metaphor: sequences of stochastic events (occurring over time) are grouped as arrays (laid out in space) that afford proportional judgment. We present classroom episodes that demonstrate how the Law of Large Numbers (many samples) can be mapped onto the classroom social space (many students) as a means of facilitating discussion and data sharing and contextualizing the content. We conclude that it is effective to embed the Law of Large Social Numbers into designs for collaborative learning of probability and statistics.*

Introduction

The mathematics domain of probability has been long regarded as challenging for students of all ages (von Mises, 1981, Hacking, 1975, 2001; Konold, 1994; Biehler, 1995; Maher, Speiser, Friel, & Konold, 1998; Gigerenzer, 1998; Liu & Thompson, 2002). At the Center for Connected Learning and Computer-Based Modeling at Northwestern University, we are creating software and computer-based activities to help students learn probability. Specifically, based on our previous research on students' challenges in understanding probability ("Connected Probability," Wilensky, 1993, 1995, 1997), we have designed a group of curricular models in the domain of probability. Our interactive models are written in NetLogo (Wilensky, 1999) and incorporate interface features that allow students to run probability experiments under different parameter settings. A strength of NetLogo is that it affords simulating many random events concurrently because it uses parallel processing (it is 'multi-agent'). Also, students can examine and even modify the code in which the models are programmed, partly because NetLogo code—its primitives and syntax—was specifically designed to be easier to read and learn as compared to other computer languages (see also Papert, 1980, on Logo). The NetLogo "models library" includes models from a range of scientific and mathematical domains. Some of these models are grouped around classroom curricular units. One of these groups is ProbLab (Abrahamson & Wilensky, 2002). This paper introduces ProbLab, focusing primarily on "Stochastic Patchwork"

and “Sample Stalagmite,” two of several models currently in ProbLab.¹ We present and analyze data from implementations in urban middle-school classrooms in which we investigated dimensions of *collaborative* activity design around essentially *individual* work with the simulations and vis-à-vis the specific content (for inherently collaborative designs, see, for example, S.A.M.P.L.E.R., Abrahamson & Wilensky, 2004).

Computer Models as Learning Environments for the Domain of Probability

We are committed to help students ground mathematical content in meaningful experiences (e.g., Wilensky, 1993, 1997; Freudenthal, 1986; Gigerenzer, 1998; Abrahamson, 2004), and our lesson plans encourage a social construction of knowledge in the classroom milieu (see Brousseau, 1997). In designing the ProbLab models, we have endeavored to make probability an approachable domain for middle- and high-school students. We submit that students’ biggest challenge with the domain of probability is not so much that the conceptual constructs per se are difficult but that the domain is difficult *as seen through the lens of traditional mathematical representations*. Specifically, probabilistic processes occur over time, and learners are challenged by the epistemological tension between, on the one hand, individual outcomes, e.g., *this* flipping of a coin or *this* sample of coin flips, and, on the other hand, phenomena as global events, e.g., the *overall* chance of the coin falling on ‘heads’ (Liu & Thompson, 2002; Hacking, 2001). In ProbLab, we are attempting to create models that allow students to move between and connect such individual (micro-) and global (macro-) outcomes (see also Abrahamson & Wilensky, 2003; Papert, 1996).

Stochastic Patchwork

An important aspect of students’ connecting to the domain of probability through working in technology-based learning environments is that students ground in probability simulations the ideas inherent both in symbolical formats of the domain and in formulae for calculating and communicating findings from probability experiments. For instance, students should *experience* the meaning of a “.7” notation in probability distribution functions or experimental outcomes. One objective of this paper is to present and discuss a type of representation that may enhance students’ bridging between, on the one hand, simple probabilistic events, e.g., flipping coins, and, on the other hand, the corresponding formal representation, e.g., an overall “.7” chance of falling on ‘heads’ (for a biased coin). Specifically, these bridging representations may help students ground an understanding of probability—what it means to say that a probabilistic mechanism has a .7 *chance* of generating a favored event—in perceptual judgments of *spatial* proportion (Resnick, 1992), i.e. seeing that .7 of an outcome array is red (see Figure 1).

¹ All ProbLab models are available for free download at <http://ccl.northwestern.edu/> see also <http://ccl.northwestern.edu/curriculum/ProbLab/> for more models, further discussion, and a complete list of our publications on design-based research, theory of learning, and equity.

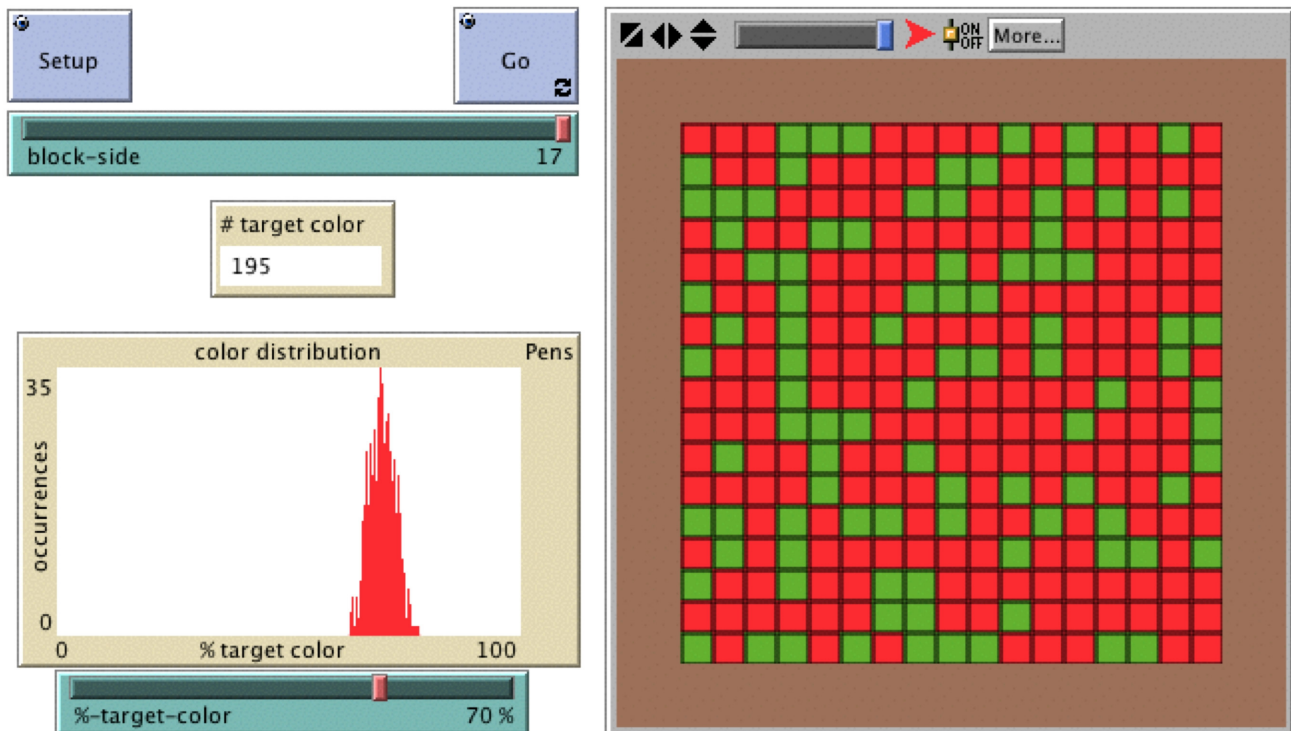


Figure 1. ProbLab: Stochastic Patchwork. Parameters are set so each square in the graphics-window “mosaic” (total of $17^2 = 289$ squares) has an independent .7 (or 70%) chance of being red on each trial. Therefore, on each trial, the mosaic has approximately .7 red squares, and after several hundred trials the histogram shows a normal distribution converging on .7.

The ProbLab model *Stochastic Patchwork* (see Figure 1, above) is a *bridging tool* (Abrahamson, 2004) between time-based probabilistic events and space-based perceptual judgments. The probabilistic element in this model is a square “coin” that has a red “side” and a green “side.” Instead of flipping this single coin many times, we flip many clones of this coin all at once. The crux of this model is that if a single coin has a .7 chance of falling on red, then the aggregate of a sufficiently-large sample of these coins that flip all at once will approximate a .7 redness, i.e. most of the time about .7 of the squares will be red. The objective of students’ interacting with the model is that they understand how the model works and explore the effect of modifying parameter settings—the size of the population and/or the bias of the coin/square—on the sample space (size and appearance) and on the dynamics of the emerging distributions.

Because the Stochastic Patchwork model simulates a probabilistic experiment, outcomes vary, yet after a sufficiently large number of successive iterations in the experiment the outcome *distribution* approximates a normal distribution converging on a .7 probability as the mean, as displayed in a histogram that is part of the model (see Figure 1, above). We have found that students as young as 10-years old working with ProbLab models interpreted experimental results using both enumeration-based strategies (counting red

and green squares in samples) and multiplicative strategies (inferring proportions in populations by eyeballing red/green ratios in samples).

Sample Stalagmite

The ProbLab model *Sample Stalagmite* helps students understand histograms of probabilistic outcome distributions by building the histograms from the outcomes themselves. The model simulates the random generation of blocks of red/green squares and their accumulation into columns according to the number of red squares in each, e.g., 0 red squares, 1 red square, 2 red squares, etc. Figure 2a, below, shows the entire combinatorial sample space of sixteen “4-blocks” (2-by-2 arrays of squares). Figure 2b, below, shows the partial combinatorial sample space of five-hundred-and-twelve “9-blocks” (3-by-3 arrays of squares). The model’s name comes from the dynamics of the visualization: the blocks descend along the columns to build a structure resembling a stalagmite (see, in Figure 2b, below, the descending 9-block marked by an ellipse).

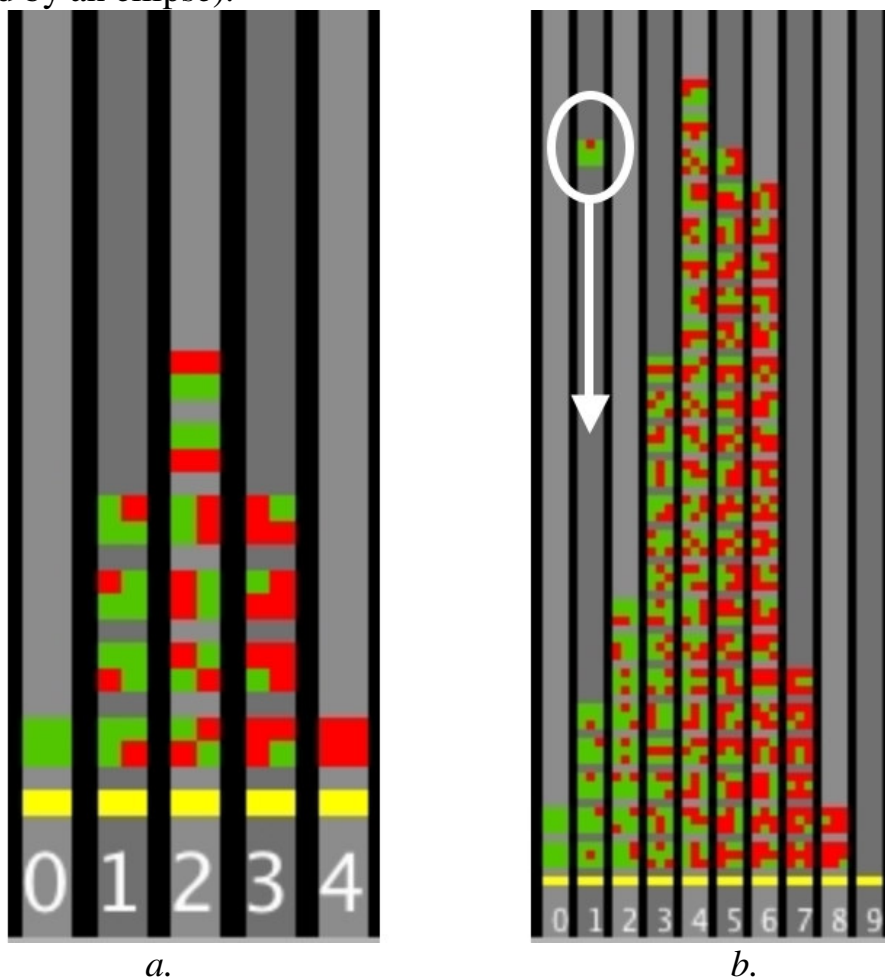


Figure 2. The NetLogo ProbLab model *Sample Stalagmite*: two fragments from the graphics window in the model’s interface under different conditions of running the probabilistic experiment.

Sample Stalagmite accompanies students’ combinatorial analysis of all possible 9-blocks (see our publications on the *combinations tower*). In this model, 9-blocks are generated randomly. That is, at every run through the procedures, one of the 512

possible arrays pops up on top of the graphics window and falls down a histogram “chute” -- its corresponding column. For instance, if there are 2 red squares in the sample, the sample will fall down the ‘2’ column (see falling 9-block in the Figure 2b, on the previous page). The model can be set either to keep duplicates or to reject them (in Figure 2 duplicates were rejected). So Sample Stalagmite takes the combinatorial space of the 9-blocks and re-positions it as a sample space. That is, each of the 512 arrays has the same chance (likelihood) of being generated on each trial.

The histogrammed combinatorial sample space is designed as a visualization bridge for students to ground a sense of the likelihood of an event in combinatorial analysis and proportional judgment. For instance (see Figure 2b), students, who are comparing the column with 9-blocks that have exactly 4 red squares with the column of 9-blocks that have exactly 3 red squares, literally *see* that the subgroup of 4-red is more numerous and taller -- it occupies more space within the histogram as compared to the subgroup of 3-red squares. This increased commonality is related directly to the fact that there are more *possible* 4-red combinations (126) as compared to 3-red combinations (84). Also, the *shape* of the stalagmite remains roughly the same whether we keep duplicates or reject them. This visual resemblance demonstrates that combinatorial analysis (theoretical probability) anticipates relative frequencies in empirical-probability experiments.

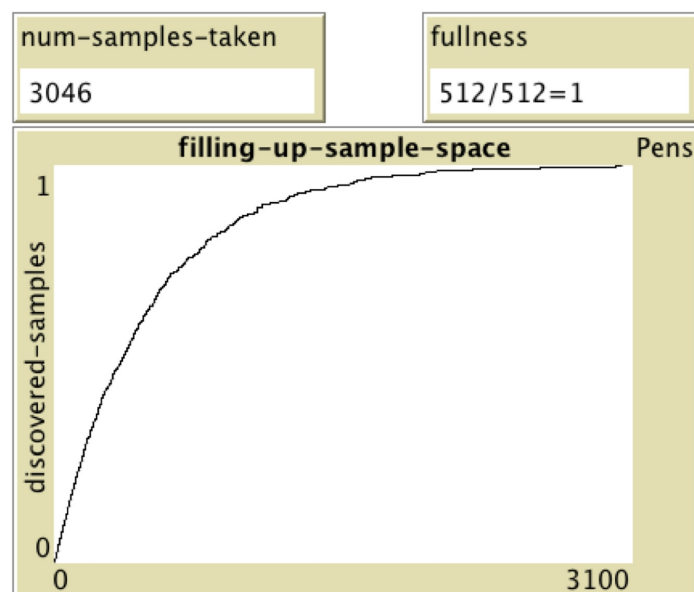


Figure 3. Fragment from NetLogo ProbLab model Sample Stalagmite.

As the simulation searches for all items of a sample space, an accompanying graph (see Figure 3, above) plots the number of discovered samples against the number of attempts. This graph invariably shapes out logarithmically. The graph is designed to support inquiry into advanced aspects of probability: Why is it that all searches for the combinatorial sample space of all 9-blocks invariably take on this shape?; Why does it take over 3000 trials to find 512 items?; Is this 6:1 ratio between trials and items significant?; Will this ratio repeat over experimental runs?: If not, why not?;

Will this ratio repeat for a sample space of size 16 (4-blocks)?; What other phenomena in the world might give rise to a graph of this shape?; How should we call this graph? These questions are nontrivial, especially when cast in terms of moving between agent and aggregate perspectives: If each specific combination is equally likely to be sampled on each turn, why is it that “the last ones are always left behind for so long?”

Classroom Research

For this particular study, we investigated a potential mapping between the distribution of experimental outcomes and the “distribution” of students in the classroom. The rationale is that just as a *single* student can take enough samples so that the shape of the outcome distribution stabilizes, so *all* students can each take fewer samples that do not stabilize unless all students pool their samples. Specifically, we explored whether such a students-to-samples mapping could be leveraged so as to stimulate inquiry-based classroom discussion.

We chose a design-based research framework (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003) so as better to investigate student early knowledge and difficulty in the domain of probability and statistics as well as to develop classroom learning supports that embrace this early knowledge and address these difficulties. Through iterated studies that began with individual students and focus groups and continued with classroom implementations, we are progressively modifying our computer-based models in response to feedback from students and teachers. Such collaboration between designers, programmers, researchers, students, and teachers, we find, is fruitful towards creating equitable and effective learning tools that may effect immediate changes in many students’ mathematical inclination and content knowledge. This report focuses on data from a single implementation along this design-research continuum.

Twenty-six 8th-grade students in a highly diverse urban school participated in an implementation of ProbLab over two weeks (half the time spent in a computer lab). Each student worked on an individual computer. Students used printed activity guides that moved from structured introductions to student-initiated experiments, and they could also modify the underlying code of the model. The researchers–facilitators and teacher moved between students for on-the-fly interviews. Each lesson included a classroom discussion. Due to the limited number of available computers, we split the students into two groups. This inadvertent staggering of the implementation proved fortuitous in that it allowed us an extra round of improvements. We videotaped all lessons both with a roaming and a classroom-spanning camera. We selected and transcribed discussion episodes to investigate how best to support students’ making sense of their experiments as they move between their personal findings and classroom pooled findings.

Results and Discussion

Analysis of classroom discussion suggests that students are intrigued and stimulated by their interaction with the models. Following, we present two data

examples. In both examples, students discuss with a facilitator outcomes from a probability experiment in Sample Stalagmite. The sampling of red or green squares had been framed as a “competition” between the two colors. Students had been asked to set the probability of red at 50%, to run the model ten times, and report their findings in terms of the ratio of red and green “wins.” In the first example, the facilitator is working only with two students, and in the second example the entire classroom is discussing their results. Note that, whereas in the first example (two students) students must *conjecture* as to the outcomes of large numbers of trials, in the second example (classroom) students’ pooled outcomes allow for a cogent *empirical* finding.

Data Example #1: “On-the-fly” interview between a facilitator and two students.

Reuven (researcher): What’s happening here?

Student1: ...more green than blue. I think the green is going to win.

R: What’s the chance that green will win?

Student1: 50–50.

R: Right. Is it possible that you do this experiment 100 times and green will win every time?

Student1: If it’s like once every f...

R: ...if it’s 50–50. Let’s say the probability is 50%. Is it possible for the green to win every time?

Students1+2: No. I don’t think it is.

R: Would it be possible for green to win if you do it...once?

Student1: Yes.

R: Would it be possible for green to win if you do it twice?

Student1: Yes.

R: Would it be possible for green to win five times?

Student1: No.

R: Wait, so what about 3 or 4? Where’s the cutoff?

Student1: It’s rare.

R: Aha! It’s rare, ok. But if I were to flip a coin a 1000 times, is it possible that it will always come out ‘heads’?

Student1: Yeah.

Student2: It will be like a miracle.

R: It will be like a miracle, but is it possible?

Student2: No.

Student1: Yes it is!

R: Why is it not possible?

Student1: Because it’s a 50–50 chance.

Student2: Well, maybe if you do it 1000 that will be your ‘50,’ and then if you do it another thousand, that will be your other 50. [Both laugh]

Data Example #2: In a classroom summary discussion, Toby had been remonstrating that “there’s something wrong with the computer.”

Toby: The probability is 50–50 [in the setting of the model], so they [the red/green outcomes] should be really close to each other, as in like it would be 4 red and 6 greens. But that’s not we had—we had like 2 red and 8 greens. So they were pretty far away from each other.

Dor (researcher): That *is* pretty far away. Were you having this all the time?

Toby: Yeah, the greens kept on winning.

Mogu: [He had worked closely with Toby] *Most* of the time.

Dor: [addressing classroom] Hands up whoever had the greens winning most of the time. [about half of the class] Ok, now hands up whoever had the reds winning. [the other half of the class] Does that make sense?

Toby: Yeah. [Dor proceeds to elicit outcomes and plot them as a distribution]

Thus, classroom discussions allowed students to share unanticipated findings from their *individual* experiments and re-interpret and reconcile these findings through the lens of *classroom* sample distributions. We concluded that probability-related simulation-based classroom activities can be designed so as to leverage and explore the randomness that is intrinsic to the content through collaborative discussion-based inquiry. When students each take a limited number of random samples, they are stimulated by their individual “wrong” outcomes to compare and compile their results as a cross-student sample-mean distribution. We have named the contextualization of the central-limit theorem in collaborative inquiry “The Law of Large *Social* Numbers.”

A promising finding is that many students appeared eager to modify the computer procedures. These students wished to individualize the appearance of their experimental environment. In particular, the students wanted to change the colors of objects on the screen. In terms of programming, this may appear as a small step, yet we believe that the act of “looking under the hood” is critical—it constitutes an easy entrance activity that allays any student apprehension of programming and creates personal precedents and a strong sense of appropriation and accomplishment.

In future studies, we will focus on: (a) understanding the conditions that best support students in linking concrete and computer-based objects; (b) the affect of printed activity guides in terms of creating shared classroom understandings and vocabulary, stimulating explorative inquiry, and facilitating opportunities for teacher attention to individual students; (c) whether more students with little if any programming experience could be drawn to modifying the computer procedures underlying the models and how such work may inform their content learning; and (d) developing a more comprehensive articulation of student understanding of the interplay of determinism (the settings of the computer model) and randomness (the specific outcomes) and how this interplay informs student cognition of the central limit theorem.

***Acknowledgement:** The design research described in this paper was supported in part by the National Science Foundation under Grant No. REC-0126227. The

opinions expressed in this paper are those of the authors and do not necessarily reflect the views of NSF.

References:

- Abrahamson, D. (2004). *Keeping meaning in proportion: The multiplication table as a case of pedagogical bridging tools*. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.
- Abrahamson, D. & Wilensky, U. (2002). *ProbLab*. The Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. <http://ccl.northwestern.edu/curriculum/ProbLab/>
- Abrahamson, D. & Wilensky, U. (2003). The quest of the bell curve: A constructionist approach to learning statistics through designing computer-based probability experiments. *Proceedings of the Third Conference of the European Society for Research in Mathematics Education*, Bellaria, Italy, Feb. 28 – March 3, 2003. <http://ccl.northwestern.edu/cm/papers/bellcurve/>
- Abrahamson, D. & Wilensky, U. (2004). S.A.M.P.L.E.R.: Collaborative interactive computer-based statistics learning environment. *Proceedings of the 10th International Congress on Mathematical Education*, Copenhagen, July 4 – 11, 2004. http://ccl.northwestern.edu/papers/Abrahamson_Wilensky_ICME10.pdf
- Biehler, R. (1995). Probabilistic thinking, statistical reasoning, and the search of causes. *Newsletter of the international study group for research on learning probability and statistics*, 8(1). (Accessed December 12, 2002). <http://seamonkey.ed.asu.edu/~behrens/teach/intstdgrp.probstat.jan95.html>
- Brousseau, G. (1997). *Theory of didactical situations in mathematics* (N. Balacheff, M. Cooper, R. Sutherland, & V. Warfield, Eds. & Trans.). Boston: Kluwer Academic Publishers.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher* 32(1), 9 – 13.
- Freudenthal, H. (1983). *Didactical phenomenology of mathematical structure*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Gigerenzer, G. (1998). Ecological intelligence: An adaptation for frequencies. In D. D. Cummins & C. Allen (Eds.), *The evolution of mind* (pp. 9 – 29). Oxford: Oxford University Press.
- Hacking, I. (1975). *The emergence of probability*. Cambridge: Cambridge University Press.
- Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge, UK: Cambridge Press.
- Konold, C. (1994). Understanding probability and statistical inference through resampling. In L. Brunelli & G. Cicchitelli (Eds.), *Proceedings of the First*

Scientific Meeting of the International Association for Statistical Education (pp. 199 – 211). Perugia, Italy: Università di Perugia.

- Liu, Y., & Thompson, P. (2002). Randomness: Rethinking the foundation of probability. In D. Mewborn, P. Sztajn, E. White, H. Wiegel, R. Bryant, and K. Nooney (Eds.), *Proceedings of the Twenty Fourth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Athens, GA, October 26-29, 2002: Vol. 3* (pp. 1331–1334). Columbus, OH: Eric Clearinghouse for Science, Mathematics, and Environmental Education.
- Maher, C.A., Speiser, R., Friel, S., & Konold, C. (1998). Learning to reason probabilistically. *Proceedings of the twentieth annual conference of the North American group for the Psychology of Mathematics Education* (pp. 82 – 87). Raleigh, NC.
- Papert, S. (1980). *Mindstorms*. NY: Basic Books.
- Papert, S. (1996). An exploration in the space of mathematics educations. *International Journal of Computers for Mathematical Learning*, 1(1), 95 –123.
- Resnick, L. B. (1992). From protoquantities to operators: Building mathematical competence on a foundation of everyday knowledge. In G. Leinhardt, R. Putnam & R. A. Hattrup (Eds.), *Analysis of arithmetic for mathematics teaching* (pp. 373 – 429). Hillsdale, NJ: Lawrence Erlbaum.
- von Mises, R. (1981). *Probability, statistics, and truth* (J. Neyman, D. Scholl, & R. Rabinowitsch, Trans.). Dover Publications. (Original work published in 1928).
- Wilensky, U. (1993). *Connected mathematics—Building concrete relationships with mathematical knowledge*. Doctoral thesis, M.I.T.
- Wilensky, U. (1995). Paradox, programming and learning probability. *Journal of Mathematical Behavior*. 14(2), 231 – 280.
- Wilensky, U. (1997). What is normal anyway?: Therapy for epistemological anxiety. *Educational Studies in Mathematics* 33(2) 171 – 202.
- Wilensky, U. (1999). *NetLogo*. Evanston, IL. Center for Connected Learning and Computer Based Modeling, Northwestern University. ccl.northwestern.edu/netlogo.

STRENGTHS AND WEAKNESSES IN STUDENTS' PROJECT WORK IN EXPLORATORY DATA ANALYSIS

Rolf Biehler, *University of Kassel, Germany*

Abstract: *The paper will point out features and shortcomings of about 60 project reports students (future teachers) submitted in a course on “Elementary Stochastics”. Findings from analysing first generation reports were used for developing a “project guide”. The analysis of the second generation reports points to further requirements for guiding students and for assessing their work.*

1. Introduction

Project work as part of the assessment in a statistics course is an opportunity for students to develop and document their statistical thinking (Wild & Pfannkuch, 1999). Projects are often suggested as part of school curricula in statistics. However, how can teachers organize and assess project work if they have never done project work in statistical data analysis themselves?

Mathematics teacher education at universities in Germany consists of three strands: mathematics, mathematics education, and pedagogy. Mathematics and mathematics education are taught in different courses but ideally their contents are related. At the University of Kassel, a course “Elementary Stochastics” is compulsory for primary and secondary teachers. The semester long course comprises 4 hours lecture and 2 hours laboratory work per week. The students can take the course with and without examination. The number of students choosing the examination option varies from 30 to 50 per semester. It is the only course on stochastics for these students. In addition, they can take a seminar on the didactics of stochastics. Since the year 2001, I restructured the course giving more emphasis on exploring data, modelling, and simulation (Biehler, 2003). The software Fathom has been used as a student tool for data analysis and stochastic simulation. Since 2002, the students are required to submit a project report as part of the assessment in the course. This has to be an individual work but they are encouraged to discuss with classmates. This type of requirement is unique in our mathematics department and the students have not many experiences in writing such reports.

As data base to be used in the projects we used a complex data set with 540 cases and about 50 variables that is based on a questionnaire concerning media use and leisure time of 540 11grade high school students: the so-called Muffins data (for details, see Biehler, 2003). The data are so rich that more than hundred different project themes have been done with this single data set. We give examples later.

The Muffins data had also been used during the course. We require the students to submit a “topic for investigation” related to this data set. The topic has to include

several subquestions that can presumably be (partially) answered by means of analysing the Muffins data set. Essentially the statistical methods to be used are measures of centre and spread, box plots, histograms and bar charts, and scatter plots. The analysis and display of distributions was required as well as a statistical group comparisons. We have not covered methods of analysing the relationship of two quantitative variables in the last two courses. However, students can work on this by reducing the problem to comparing distributions of a second variable for different levels of a first explanatory variable. They recode the first numeric variable into several categories such as *very low*, *low*, *medium*, *high*, and *very high* or use a separation into four equal-sized groups according to the quartiles. Students' can compare the distributions of the second numeric variable for the 4 or 5 levels of the recoded first variable.

In the first generation, the students were prepared for the project work by the following components. They had got homework assignments with smaller data analysis tasks, the lecture and the laboratory sessions included longer examples of data analysis, which were orally presented and discussed. The students were then asked to do an analysis, write a report, and submit a first draft of their project report. We gave detailed feed-back on the first draft and required them to submit a second final report. From an analysis of the short-comings of the first generation final reports, we developed and handed out a guideline for doing statistical projects and for report writing, briefly called the *project guide*. Equipped with this improved material, students of the second generation had to submit only one report that counted as the final one. The length of the reports should be 8 pages of text plus graphs and tables.

The second generation reports were analyzed in a master thesis and selectively compared to the first generation reports(Heckl, 2004). As a result new recommendations and tools for writing project reports will be developed for a third generation. We got 30 first generation reports and 33 second generation reports. A project could involve one or two students. Currently, 50 third generation reports are in the process of being assessed.

2. Topics and questions from the students' projects

The following titles are examples of students "topics for investigation":

- 1) "The relation between body mass index and use of TV"
- 2) "The influence of having a job on leisure-time activities"
- 3) "Sex differences in sleeping times of females and males"
- 4) "Computer use and homework"
- 5) "Reading"
- 6) "Jobbing"
- 7) "Sports Club"
- 8) "Newspaper reading"
- 9) "Shopping and strolling through towns"
- 10) "Sex differences in leisure time activities"

Sometimes the title itself contains already a “research question” such as in the first three ones. Many titles, however, just contain a “research domain”. Students structure such a “research domain” in very different ways. For instance, one student broke “Jobbing” (6) into the following partial questions: How many hours do students work for a job? What is the influence of jobbing on living in affluence of the students? (Are women disadvantaged? Do students finance expensive devices and equipment by themselves? Is jobbing done for ‘la dolce vita’?) Where do jobbing students get the time for jobbing from? (Less helping at home? Do they do less homework with the result and worse grades? What about sleeping time and time for music and sports?). This is an excellent example where a question was developed from an authentic subject matter motivation and interest. The richness of the data set allowed different paths through the same topic. For example, the student who chose project (2) discussed for instance, how jobbing affected leisure-time activities that cost money. The data included two variables on the frequency of going into discos and into pubs respectively. Differences in the distribution of these variables between those who had a job and those who don’t were analyzed. But no information was available on how much money the students had at their disposal and how they spend it. This limitation of the data was critically reflected when the student summarized her findings.

We consider these two as good examples. At the other extreme, the student who took “Newspaper reading”(8) just made 8 group comparisons of males and females with regard to the eight variables that are part of the data set and correspond to interest in different columns of a newspaper (politics, local news, sports etc.). This is an example of a project type where students just picked out a subset of variables on which they schematically did group comparisons. Subject matter aspects are not excluded but were not the driving force and were treated superficially.

3. Structure of report writing and statistical thinking

Wild and Pfannkuch (1999) developed a process model for statistical thinking. Their general framework is the PPDAC cycle: *problem* → *plan* → *data* → *analysis* → *conclusions*. The authors have refined this framework in various ways. This framework is helpful for structuring the *process* of data analysis as well as for structuring a project report. The projects our students had to do differ in one important aspect from the usual PPDAC cycle: The Muffins data were already given. Instead of planning the collection of new data that fit to the question, students were asked to select from the available variables to fit their questions and - if applicable - to point out in which sense the available data were not sufficient or adequate for answering their questions.

We suggested that students structure their report into 3 parts: (1) Introduction (2) Analysis (3) Summary and Conclusions.

The (1) *Introduction* is to describe the development from problem to plan of analysis including the selection of variables. The students were required to formulate hypotheses and qualitative expectations. The (2) *Analysis* should not only present re-

sults but also let the reader participate in the process of interactive data analysis. Here we follow the recommendation of Wolf (1989) that a report especially in EDA should contain possible hypotheses, alternative analyses and alternative conclusions according to a principle of disclosure and transparency that is important for such a type of exploratory analysis. A justification of results cannot be completely separated from the process how they were gained. In this sense our students were encouraged to report on discoveries they had not expected and on ideas they came across during the process of analysis. This does not mean that they should present a protocol of the sequence of their actions but already something that was filtered through a process of data synthesis. The (3) *Summary and Conclusions* should contain a more concise summary of the most important findings.

Both generations of project reports we got had severe limitations with regard to introduction and summary. There were extreme cases without any introduction and summary. For instance, a student chose the topic “What is the difference between males and females with regard to time watching TV”. In the data set there were about 5 variables related to TV. The student chose one after the other and did a group comparison between males and females for all 5 variables and ended with the last comparison without any synthesis of findings. In the “project guide” we pointed out how important an introduction is for attracting a reader and for elaborating a question. We got improved results in the second generation. Whereas 83% of the first generation reports contained an introduction, it was a full 100% in the second generation. Also the quality of the introductions improved as the following table is pointing out.

	<i>First generation</i>	<i>Second generation</i>
<i>Proportion of reports with an introduction, where...</i>	%	%
topic is explained and discussed	28	55
selection of variables and research plan are presented	32	76

Table 1 Features of introductions of the reports

Nevertheless, we do not think that we have yet reached a satisfactory level of quality.

Writing a summary was also a problem for our students. I call the two extremes of summaries we got “the journalist summary” and “everything is repeated summary”. In the first extreme only qualitative results were reported such as “girls tend to help more with the housework than boys do”. At the other extreme students just repeated all details of all graphs without daring to select and summarize most important information.

Our analysis underlined that report writing as such introduced a lot of new requirements from our students that are different from the competence needed for *doing*

statistical data analysis. Jambu (1991) coined the term “data synthesis” for the process, where the results of an exploration have to be ordered, compared, assessed according to importance, refined and presented to a potential audience in a convincing way. Data synthesis involves preparing an act of communication that may need specific means of communication, such as graphs and convincing arguments that anticipate possible criticism. This partly corresponds to the last step of Wild & Pfannkuch’s PPDAC cycle: Drawing conclusions.

Being able to concentration on the most essential and a good “style” of presentation is important but on the other hand the meagre contents of journalist summaries have to be avoided.

4. Weaknesses in students’ project reports and recommendations for improvements

The guide for doing projects in the second generation was set up after we had analyzed the first generation of reports (Heckl, 2004). Our “project guide” contains the following “attention topics (1) Graphs and Tables (2) Interactive and experimental working style, group comparison, style of writing. The following aspects were listed under “1 Graphs and Tables”

- 1) *Setting up multiple graphs*: use same scales, use adequate juxtapositions of graphs, number graphs for referencing in the text, do not add too many summary values in a graph, use numerical summary tables in parallel to graphs);
- 2) *Reading off information from the box plot*: checking whether outliers in the box plot may be just indications of long tails or whether there are specific causes, check for exact relative frequencies (middle box has only approximately 50% of the data points), use the concept of density for describing structure in box plots);
- 3) *Histograms*: use relative frequency for group comparisons, be aware that intervals are open at the right end, and choose one or more reasonable bin widths for class intervals);
- 4) *Percentile plot*: compare to other distribution displays and look for specificities that can be seen in this plot; in comparison of distributions: check whether a statistical variable is generally statistically larger than a second variable).
- 5) *Group comparisons of two categorical variables*: choice of adequate bar charts with proportions relative to the subgroups not to the whole group, more adequate summaries than just comparing bar by bar);
- 6) *Multiple graphs*: Awareness of the relative strengths and weaknesses of the available graphs such as the histogram, the box plot, and the percentile plot.

Under the heading of “2 Working style etc.”, the following problems are discussed: How to formulate a statistical problem and to plan an analysis, how to write an interesting introduction, how to choose better summary statistics for better communication, how to work interactively (being inspired from one analysis and then go-

ing a step further and raise new questions), comparison of groups, distinguishing description and interpretation, avoid misinterpretations such as too easy causal explanations instead of just distribution differences, synthesis of results, understandable writing.

Whereas the problems related to graphs and tables (1, above) did largely improve in the second generation of reports, the problems related to working style (2, above) largely remained. Although the written guide was relatively short other parts of the course discussed these problems and took into account the results we have reported on elsewhere (Biehler, 1997, 2001).

From the analysis of the second generation reports by Heckl (2004) the following recommendations can be made.

(1) Students need to study a *prototypical project report*, where the underlying general aspects of the concrete report are made explicit. We think that a two column presentation is valuable, one column with the text of the report and a second column with comments that point out for which general principle the current text is an example. We had also put good students' reports from an earlier generation on the web, but many students seem not to grasp intuitively the quality features of such reports.

(2) We think of "analysis of a single distribution" and "group comparison" as integrated and networked competencies. Students find the integration of the individual graphs and methods into a cognitive competence module difficult. A detailed guidance and prototype for each of these two basic competence modules is necessary, containing check lists and hints how to relate graphs etc.

(3) Summary and introduction writing should be guided by showing a variety of solutions for these parts of the report in order to communicate a "feeling" of quality. Formal elements of writing a scientific report (on whatever subject) are often not familiar.

(4) Limitations of data analysis and the risk of haphazard conclusions should be pointed out. Carefully weighing and presenting evidence is a virtue of a data analyst that has to be cultivated. On the other hand an exploratory spirit and attitude in collecting or selecting data and in interactive data analysis has also to be cultivated. Students have to be made aware of the need of this role change.

Students differ largely with regard to this latter problem area and it is most unclear, how we can get improvements in this respect.

5. A prototypical module: Analysis of statistical distributions

In this section, I try to give very concrete details with regard to the competence module "analysis of one distribution" we have in mind.

From the project reports we learned that even analysing univariate data was more difficult for students than we expected. Of course, all students can produce the standard displays such as histograms and box plots and rephrase what can be easily

seen in these displays, but this is not very satisfactory. We have still to clarify what we should consider as good or excellent analyses of univariate data and how we can grade students' written interpretations.

Let us take the variable *Time_Homework* from the Muffins data set as an example: The 11graders (16 or 17 years old students) had to tell how many hours per week they devote to doing homework. What are components of student competencies in this context?

We can pose very closed questions such as: What is the mean? What is the proportion of students working more than x hours per week? From the perspective of EDA we would ask: How are the values distributed? And we would have some expectations and should be open to unexpected behaviour of the data.

(A) "Estimating" a distribution

This means developing expectations about the distribution of a variable by means of using context knowledge. For instance, we could reason as follows:

(a) 0.5 up to 2 hours on average for each of the five workdays seems to a reasonable range for doing homework. This means that most of the data will vary between 2.5 and 10 hours per week. I expect a small proportion outside this interval. Maybe 5% less than 2.5 and maybe 10% more than 10 hours (I know that there is always a subgroup of hard-working students. An average or a typical value is difficult to estimate, if it is 1.5 hours per day, then we get 7.5 as a mean for the whole week.

(b) I expect the distribution to be skewed to the right (from knowledge of other leisure time variables), and I expect popular values at multiples of 5 because student tend to round to the nearest 5.

Whereas the aspects listed under (a) refer to context knowledge, (b) shows the statistical expert who uses analogies to other statistical variables for developing expectations. Estimating a distribution activates context knowledge and sets an expectation, a perspective from which the data are seen and to which actual data can be compared. With a context-based expectation an external reference point is created that may help students interpret their findings.

(B) Combining information from various distribution graphs

The tools students had at their disposal are the following graphs: histogram (with adaptable bin width), dot plot, box plot, percentile plot. It was possible to enhance graphs by statistical summary values that are definable by the formula editor of Fathom, such as median and mean. Statistical (summary) values can also be displayed in a summary table. I try illustrating what we recommend by a detailed prototypical example.

(1) Use basic displays.

- a) Use the box plot and the default histogram as a starting point; use the same axes and arrange the 2 graphs exactly vertically in order to improve

comparisons; add a summary table with the values of the box plot and the “count” of the variable, which shows the number of non-missing values for this variable.

- b) Check differences and commonalities between box plot and histogram: Do they indicate the same type of distribution (symmetrical, skewed)? Is the density information in the box plot and the histogram compatible with each other? Does the histogram add further structure such as bimodality or popular values? How do box plot outliers show up in the histogram?
- c) Add the mean to the summary table and if useful also to the plots and try to explain large differences between median and mean from properties of the distribution

(2) *Refining frequency information beyond standard displays.* If the data contain ties or the sample size is not very large: Check how well the overall frequency information of the box plot is true (should always be approximately 25%). Use more fine-grained histograms and percentile plots and check whether they reveal more interesting structure to you.

(3) *Expected and unexpected features.* Compare the results to your initial expectations. What did you expect, what is unexpected? Do additional analyses according to your expectations if necessary (For instance calculate the proportions of those with homework above 10 hours that you expected to be about 10% before starting the data analysis). Are there unexpected interesting features in the distribution that deserve further study? If possible try to find out more about subgroups and specific causes that might have influenced the variable displayed in the histogram.

(4) *Summary and interpretation.*

- a) Select those distributional aspects that you find most important. Do you think that communicating selected frequency information is important? Will a recoding of the data into 3 or 4 categories give interesting information?
- b) Relate your results to external reference information. Try to give a deep contextual interpretation of the results. You must not just rephrase results by using words from the context.

The result of (1a) is shown in Fig. 1. Matching the axes and exact juxtaposition was missing in many first and second generation reports. Students also often tried to estimate the summary values from the display, or to get the numerical values from plotting all the summary values into one display hereby making it overcrowded. (1b): box plot and histogram show a distribution skewed to the right. The density information in both displays is compatible. The histogram shows some bimodality with a peak in the interval [10,12). Whether the bimodality is an artefact of the data collection or of this specific data display or an indication of two discernable subgroups could be something for further investigations. The outliers show up similarly in both

displays. The outlying values seem to indicate a long tail rather than being largely separated from the other data.

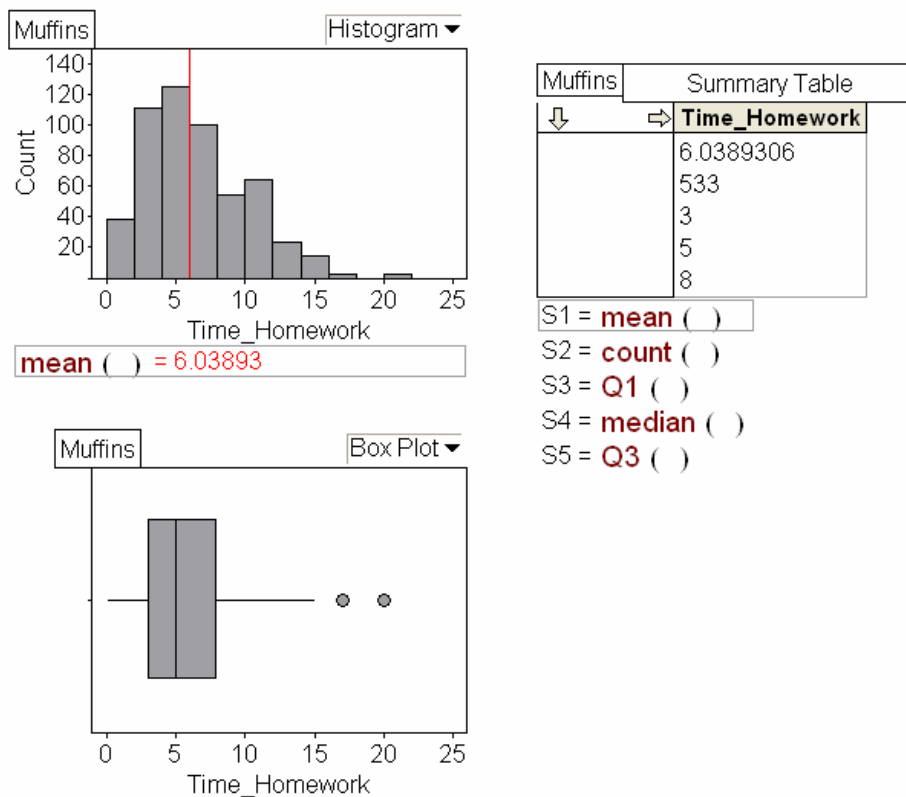


Fig. 1 Graphs from a distribution analysis using Fathom

(1c) The mean is 1 hour higher than the median. This corresponds well to the graphical skewness we can notice.

(2) Fine-grained histogram, histograms with different starting values and percentile plot do not provide more interesting structure (the plots are omitted here). We exemplarily check the box plot frequencies and get 61 % for the “middle half” (the grey box including the border points). In Fathom we can use a specific formula or a formula with placeholders. The latter will always display correct frequencies when we change the variable in the summary table.

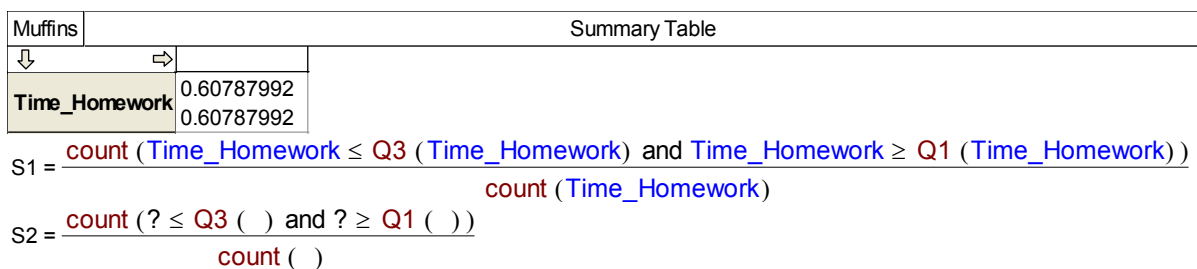


Fig. 2 Additional frequency information about (data as in Fig. 1)

This recommendation was introduced because students tend to “translate” the statement: The quartiles are 3 and 8 into “the middle 50% of the values are lying between 3 and 8.”

(3) Compared to our expectation the centre is lower, there are even students who do no homework at all, but also see a much larger upward tail than was expected, with values up to 20. The percentage of those who work less than 2.5 hours is 14.4% that is much more than expected, but we see from the graphs that many work 2 hours, and the percentage for less than 2 hours per week is 7%. 8.5% of the students do work more than 10 hours.

(4a,b) In addition to the box plot the following categorization can be intelligible. Very low (less than 2.5), medium low ([2.5, 5]), medium ((5,10]), high (above 10). This is related to an external frame of reference (How much per day on average) and this categorization gives the above result. Due to the unequal bin widths, the shape of the distribution changed.

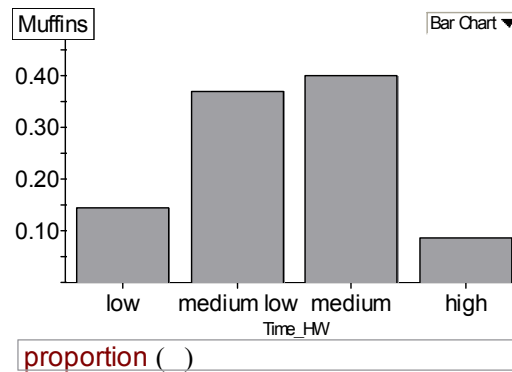


Fig. 3 Additional display for the frequency information (data as in Fig. 1)

6. Summary and future plans

Our work on analysing students’ projects, developing a new project guide, and developing an assessment scheme is still in progress. I showed some details and examples in this paper. A very complex module is the module on group comparison that will be discussed elsewhere. Moreover, style of interpretation and of relating the context to the statistical question is highly variable between the students. Another aspect is the formal structure, the sequence of questions in a report. In a next step, we will try to develop a representation for this aspect. We plan to do a more complete analysis by means of using the software *atlas-ti*, which will help us to categorize parts of the project reports and make more easy comparisons between them.

7. References

Software

FATHOM™ <http://www.keypress.com/fathom/>
or German version: <http://www.mathematik.uni-kassel.de/~fathom>

Other References

- Biehler, R.: 1997, 'Students' difficulties in practising 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*. Voorburg: International Statistical Institute [also <http://www.stat.auckland.ac.nz/~iase/publications/8/14.Biehler.pdf>], pp. 169-190
- Biehler, R.: 2001, 'Statistische Kompetenz von Schülerinnen und Schülern - Konzepte und Ergebnisse empirischer Studien am Beispiel des Vergleichens empirischer Verteilungen.' in M. Borovcnik, J. Engel & D. Wickmann (eds.), *Anregungen zum Stochastikunterricht*. Hildesheim: Franzbecker, pp. 97 – 114
- Biehler, R. (2003). *Interrelated learning and working environments for supporting the use of computer tools in introductory courses*. Paper presented at the IASE Satellite Conference on Teaching Statistics and the Internet (CD-ROM Proceedings)[also: <http://www.stat.auckland.ac.nz/~iase/publications/6/Biehler.pdf>].
- Heckl, R.: 2004, *Die Bewertung von Projektarbeiten zur Explorativen Datenanalyse in der schulischen und universitären Ausbildung*. Unpublished Zulassungsarbeit Erste Staatsprüfung, University of Kassel, Kassel.
- Jambu, M.: 1991, *Exploratory and Multivariate Data Analysis*. London: Academic Press.
- Wild, C. J., & Pfannkuch, M.: 1999, 'Statistical thinking in empirical enquiry', *International Statistical Review* 67(3), 223-265.
- Wolf, H. P.: 1989, *Grundprobleme der EDA - Literate EDA als Antwort auf Kommunikationsprobleme einer explorativen Datenanalyse*. Münster: Lit.

RANDOMNESS AND LEGO ROBOTS

Cerulli Michele, Chiochiariello Augusto, Lemut Enrica

Istituto per le Tecnologie Didattiche- CNR di Genova

Abstract

This paper reports on a long term experiment concerning the introduction of 7th grade pupils to the concept of randomness. Pupils are involved in activities with Lego robots, and in the joint enterprise of writing an Encyclopaedia. The main lines of the experiment are provided, together with experimental data, highlighting how some specific elements of the chosen educational approach influenced the evolution of pupils' mastery of the concept of randomness.

1. Introduction

The research we are presenting has been developed in the framework of the Weblabs¹ project, which focuses on “new ways of representing and expressing mathematical and scientific knowledge in european communities of young learners”. The teams involved in the project focused on a variety of scientific concepts, developing and testing specific educational approaches based on ad hoc designed technological tools; in particular, our team focused on the concept of “randomness”.

The tools used are based on the programming environment ToonTalk (Kahn 2004), and on a computer supported collaborative environment. Moreover, our team was in charge of designing and testing Lego RCX robots, interpreted as advanced technological artefacts embedding knowledge concerning randomness. In a sense, a key assumption is that technological artefacts, such as Lego robots and ToonTalk programs, can be considered as reifications of randomness-related concepts.

In this paper we focus, and discuss, on two main findings concerning the influence of the educational approach employed by us on how pupils learnt about randomness. The first one regards the students' capability to substitute each different random generator in a given physical device; the second one concerns the students' capability to differentiate random from not-random sub-elements in a system.

2. Theoretical framework

What is randomness? What is a random phenomenon? Given a phenomenon how can we judge if it is random or not?

These questions are still open, in the sense that there is not yet a universally accepted definition of randomness. In fact mathematical probability is a quite recent subject, and historians chose 1654 as a convenient landmark for its birth, due to the contents of the correspondence of Pascal and Fermat regarding games of chance (Volchan, 2002). Furthermore its first universally accepted axiomatisation was proposed by Kolmogorov in 1933. Humans have however been coping with randomness for thousands of years, for instance in games of chance, thus it is only its mathematical formalizations that are relatively new. The peculiarity of mathematical formalizations of randomness is that they are based either on common sense, or on key ideas derived

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/>).

from various scientific contexts. In fact we can find interpretations, and related attempts at formalizations, of the word *random* as: *unpredictable, lawless, incomputable, uncompressible, not deterministic*, etc. Any of such characterisation can be ascribed to the idea of randomness, and contributed to define its key aspects, as shown by the historical evolution of the definitions of randomness.

According to this brief historical sketch, it is not surprising that the learning of the concept of randomness (and the related concept of probability) may be difficult, as witnessed by related research literature (Pratt 1998, Wilensky 1993, and Truran 2001). In particular we may focus on the following key educational issues.

Issue 1. A variety of meanings derived from a variety of experiences

The learning of the concept of randomness may be hindered by contrasting views derived from different experiences or from socio-cultural biases. Actually Nisbett (1983) points out the sensitivity of children's response to the situation, and Pratt remarks that "at a low grain size, we see notions of randomness as disconnected pieces of knowledge, with different resources generated by changes in settings" (Pratt, 1998). This suggests a need of reflecting on different experiences in order to connect them and build an integrated idea of randomness.

Issue 2. Too much emphasis on determinism can be counter productive in schools

Fischbein's research highlighted how school's emphasis on causality and determinism may have a counter productive result (Fischbein 1975, p.73):

"This is why the intuition of chance remains outside of intellectual development, and does not benefit sufficiently from the development of operational schemas of thought, which instead are harnessed solely to the services of deductive reasoning".

In other words, we can argue that there is a need to put emphasis on indeterminism and randomness, in order to develop intuitions of chance. Moreover, Fischbein suggests: "in order to create new correct probabilistic intuitions the learner must be actively involved in a process of performing chance experiments, of guessing outcomes and evaluating chances, of confronting individual and mass results, *a priori* calculated predictions, etc. New correct and powerful probabilistic intuitions cannot be produced by merely practicing probabilistic formulae. The same holds for geometry and for every branch of mathematics." (Fischbein, 1983, p.12).

Issue 3. Needs of theoretical reflection

But, even if certain ad hoc designed experiences may help the development of intuitions, this does not guarantee the development of underlying mathematical ideas and structures, as commented by Pratt (1998, p. 44):

"[...] schools might adopt a pedagogy in which children play games in order to experience randomness and build on this informal knowledge, though as I observed in earlier sections such approaches do not necessarily offer a very high chance that the children will attend to the mathematical structures within the game."

Konold (Konold, 1995, pg. 209) argues that simulations offer us a way of testing our

theories, not replacing them, and that theories should remain the primary focus:

“My own belief is that this approach has a chance of leaving untouched the informal notions students bring into the classroom. The approach I have used is to encourage students to articulate their informal theories, to make predictions from them, and to use the results of simulation to motivate alternative explanations.”

Konold argues (ibid, p.184) also that:

“Typically, people dichotomize, seeing phenomena as “wholly random” or as deterministic. The kinds of constructions made by the interviewees, the negotiation of meaning for randomness, probability and distributions, are the kinds of bridges necessary to a less dichotomized view.”

These observations suggested to us the need to develop an educational approach based also on pupils' social construction of a knowledge concerning randomness shared by the class. We argue that a useful way towards this goal is to guide and help the pupils, individually and/or as a group, in verbalizing and communicating their evolving knowledge in some steps of the teaching-learning process.

Issue 4 The mediating role of technologies

A wide body of literature exists concerning the mediation role of technologies in relation to the learning of mathematical concepts (Noss & Hoyles, 1996; Bottino 2001). Research such as that conducted by Pratt (2002) and by Wilensky (1993), suggests that microworlds can be fruitfully employed as means for achieving educational goals related to probability. Moreover, Papert suggests a way of empowering the idea of probability by setting up activities that include sample space manipulation, and employing probability (and randomness) as a strategy for problem solving in contexts involving computers and programmable robots (Papert 2000).

Our research is based on the idea of using different microworlds as sources of a variety of meanings that must be integrated in order to build the concept of randomness, crucial for understanding probability. We believe that such meanings can be integrated by setting up activities where different microworlds can be compared and connected by focusing on their random aspects. In particular, we use two specific microworlds: the first one is physical and tangible (Lego RCX), the other one (ToonTalk) is virtual and embedded in the computer.

3. The Activity Sequence implemented and experienced

3.1. Basic hypotheses

Coherently with the presented theoretical framework, we chose some working hypotheses, functional to the aims of the research. We assumed the importance of:

- developing an investigative atmosphere, giving the students situations to explore;
- focusing pupils' attention and reflection on the distinction between random phenomena and non-random phenomena;
- fostering pupils' capability to assume different standpoints in order to observe, or reflect upon, a given random-related phenomenon, object, or fact;
- pupils' involvement in a variety of experiences involving different kinds of microworlds (in a wide sense), in order to characterize the concept of randomness;

- setting up comparison activities between the different experiences, stressing analogies and differences;

One of the educational aims is that each pupil builds a possible unifying model, to be used to describe different random phenomena.

3.2. The design

The designed approach to randomness relies on the exploration of some *key concepts* (eg. *predictability, unpredictability, fairness, unfairness, determinism, indeterminism*, etc.), and of some *key properties* of random phenomena (eg. the properties of random walks, the independence of events from their history, etc.). The selected concepts and aspects of randomness are explored in three main phases:

Randomness Small Talks: a collection and analysis of sentences, talks, previous experiences made by the students, directly or indirectly, where the random concept emerges in some way.

Phenomenological approach to randomness: based on the manipulation and reflection on the nature and functioning of ad hoc designed RCX LEGO robots.

Toward mathematization: some ad hoc designed computer microworlds, based on ToonTalk, are used to introduce a formal language and mathematical formalization.

In each phase, pupils are required to write individual and collective reports on the activities. In particular the class is engaged in the joint enterprise of building a shared *Encyclopaedia of randomness*. The items of the produced encyclopaedia (and their contents) are derived from the class experiences and from individual and group reports, and are meant to represent the shared culture of the class (Cerulli & Mariotti, 2003). The general methodology is that of negotiating the contents of the encyclopaedia by means of class mathematical discussions (Bartolini Bussi, 1996). Items in the Encyclopaedia are thought of as evolving entities, and in practice they are revised and updated periodically by the class along with the experiments.

3.3. The experimentsl setting

The experiment is a long term one (2 years, the second of which is in progress), and involves pupils from different european sites participating in several activities for each of the described phases. In this paper we deal only with some activities of the first two phases, which took place in the first year, and concentrate on the data concerning a group of pupils situated in Italy.

We worked with a class of 23 pupils (7th grade, 12-13 year old) in a compulsory school near Milan (Italy). The experiment has been included in the science and maths curriculum of this class, as set out by local autonomy rules on experimentsl activity. The class was provided with a portable computer and internet connection, and could occasionally also use 10 computers in the computer laboratory of the school. In total 19 sessions were set up, 13 of which lasted 110 minutes, the remaining ones varied from 25 to 55 minutes, and the last 6 were dedicated to the second phase of the activity sequence. Such a phase consists of several activities involving Lego robots. For each of the 3 employed robots, we set up a session of 110 minutes with practical tasks involving the robot, and a 110 minute session consisting of a class discussion

aiming at updating the Randomness Encyclopaedia.

3.4. The context and the submitted tasks

3.4.1. First Phase

In the first phase (called “Randomness Small Talks”) pupils are asked to present examples of events related to randomness (Fig.1), and to discuss their random or non-random nature (Fig.1, Task B). Similar activities are then submitted concerning examples of predictable and unpredictable events, and concerning a study of games, proposed by pupils, in terms of randomness and predictability.

Task A: Randomness. Have you ever heard phrases containing the expressions "by chance" or "randomly"? Write these phrases..

Task B: Randomness. We need to agree on the meanings we attribute to the adjectives "random" (or "by chance") and "not random" (or "not by chance")². Write an individual text describing a "random" situation and a "not random" one, use the following schema:

WRITE: examples of "random" situations³
 INCLUDE: drawings and/or pictures that you find relevant
 EXPLAIN: why you think such situations are random ones
 WRITE: examples of "non random" situations

In class we are going to discuss your texts in order to reach shared meanings for the expressions "random" and "not random".

Fig. 1: The first two tasks submitted to pupils in order to introduce the theme of randomness and to distinguish between random and non-random events. In the Italian text, we use the expressions “*per caso*” and “*a caso*”, respectively for *by chance* and *randomly*.

The first phase ends with a final task in which pupils are required to write a collective class report concerning the meanings of the words “random”, “non-random”, “predictable”, and “unpredictable”. They produce the first items of the class Randomness Encyclopaedia, where the contents of the items are socially negotiated and are then structured according to a given template (Fig.2).

Title of encyclopaedia item:
Meanings:
Examples:
Synonyms and contraries:
Related Weblabspedia items:
Curiosities / Anecdotes / Miscellanea / History:

Fig. 2: Template for Encyclopaedia items.

3.4.2. Second phase

The employed robots have been built by us on an ad hoc basis, and have different levels of transparency, manipulability, and interactivity, as far as their random components are concerned. The first robot that we presented to pupils, the *ShakerBot*, can be driven by a user by means of a special device, the shaker: when the device is shaken, the robot executes a walk, which can be random or not, depending on how

² The Italian word *casualmente* means either *random* or *by chance*, depending on the context.

³ The Italian *situazioni*, which we translated with *situations*, stands also for *contexts* and for *facts*.

the user moves the shaker. In this case the source of randomness consists of the user together with the shaker. In the second robot, the *Drunk Bot*, the source of randomness consists of a mechanical device that is part of the robot, as we will better describe below. In these two robots the devices that are the source of randomness can be easily observed, manipulated and modified, thanks to the properties of their LEGO components. The last robot that we used, the *Sweeper Bot*, is programmed to move randomly by means of a standard computer random function, which is its source of randomness. In this case its random component is hidden, it is a black box, but it can be used to study the properties of the random walks it produces.

In this paper we focus only on the activities that involved the *Drunk Bot* (Fig. 3). This robot is a vehicle that can execute only two kinds of movements: step forward, and step backward. A special component of the robot, is a random generator (that we called “Roller”), consisting basically of a slide, a pin, a marble, and two sensors (Fig. 3). At each step, the robot “decides” to move backward or forward, according to the sensor hit by the marble in the roller device. In a sense, the robot simulates the walk of a drunk man who is not able to decide whether to go forward or backward. The resulting movement is a one dimensional random walk.

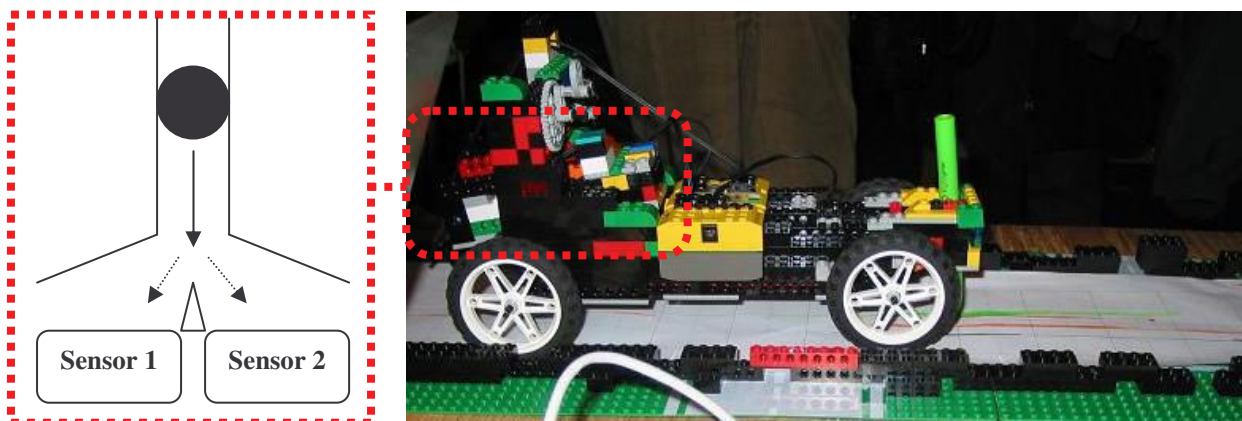


Fig. 3: On the right, the Drunk Bot is free to move on a lane, leaving a coloured trace thanks to a pen. The Roller device consists of a transparent component of the robot which is explained by the left picture. A marble slides down and hits a pin, then it may go left or right (randomly), thus hitting sensor 1 or sensor 2. The Drunk Bot moves a step backward or forward according to the hit sensor.

The task proposed to the pupils requires them to produce and justify conjectures concerning the positions of the robot, after a while. For example: “*where is it going to be?*”; “*Is it going to be close to, or far from the starting point?*”: “*Does it move forward or backward more?*”. The task is developed in the form of class observations and discussion; the focus of the discussion is guided by the teacher by means of posing questions.

At the end of the second phase, a final Randomness Small Talks is set up, in which pupils are explicitly required to analyse the Lego robots, and to classify them in terms of being random or not random, predictable or unpredictable. The conclusion of the activity is the updating of the Randomness Encyclopaedia. In particular, the teacher brings into class a poster containing all the encyclopaedia items previously developed by the class, and containing photographs of each Lego robot. Pupils are asked to

update the poster, indicating, for each LEGO robot, if it can be included as an example for either the “random” or the “non-random” encyclopaedia item.

4. Results and discussion

In the following we present and discuss some results gathered from the data collected during the Initial Small Talks, and highlighting some aspects of pupils’ knowledge related to randomness that evolved throughout the experiments. Consequently we will show evidence of this evolution, by presenting data from the Final Small Talks, and highlighting how the employed educational approach fostered such evolution.

4.1 Some indications from the initial Randomness Small Talks

In all the examples proposed by pupils the main actor is a human one, and in most of the cases such an actor is the pupil herself/himself. We find for instance pupils proposing examples of random situations such as “*I chose a jacket randomly. (without thinking)*” and “*I found a coin by chance (luck)*”. In such examples, the pupil is a constituent part of the considered random phenomenon. In these cases it may be difficult, for the pupil, to assume an external standpoint, which could result in a difficulty in understanding the complexity of random phenomena. As a consequence we believe that there is a need to consider situations where the pupils are not the main actors of the involved random phenomena. A situation of this kind is suggested by the following example proposed by Ciufciuf (one of our pupils): “We are chosen randomly to be examined”. In this case the main actor is “the teacher” who participates in the “random” phenomenon of choosing the pupil to be examined. The difficulties of changing stand point is demonstrated by the following excerpt of a text written by a pupil (Vale) reporting a class discussion concerning the random nature of the considered situation:

“Ciufciuf said that for us pupils the sentence could be random, because we don’t know who will be chosen for interrogation, while for the teacher it is not random because she can decide who she is going to interrogate. [...] Not everyone was convinced so the teacher asked us to elaborate with other examples...”

Ciufciuf attempts to analyse the phenomenon by assuming two different standpoints, but this attitude remains isolated and the rest of the class does not follow his position. Here we observe that at each step of the proposed activity sequence, pupils are required to discuss the nature of the considered phenomena, trying to reach a shared position in terms of classifying the phenomena as random or non random.

4.2 Some indications from the final Randomness Small Talks:

In this part of our experiment the employed Lego RCX robots were pre-built tools, whose peculiar characteristic was their “transparency” for the users. This transparency allowed pupils to investigate the different components of the robots, and their specific functions, providing a rich source for reflecting on randomness, as shown by the examples provided in what follows.

4.2.1. *Is the Drunk Bot random or not?*

During the final Randomness Small Talks, pupils are asked to discuss the random/non-random nature of the Lego robots, in order to reach an agreement to be

expressed in the form of an encyclopaedia item. In particular they discuss the random/non-random nature of the Drunk Bot. In the following we are going to analyse some key steps of that class discussion.

4.2.2. Step 1 - The Drunk Bot is not random!

The episode begins with the teacher asking pupils to express their opinion (numbers indicate the chronological order of the excerpts within the class discussion):

1. T: What about the drunk one? (meaning “is it random or not?”)

5. C1: so, the drunk one, from our point of view moves randomly, but from its point of view it does not...does not go randomly...

6. Many voices, we can hear many different opinions!

First of all, we observe that C1 seems to be able to judge the situation changing standpoint. In fact she talks both of “our point of view” and “the robot’s point of view”. Such a shift of standpoint, enables her to question the nature of the drunk bot assuming a position which starts a rich and meaningful discussion among pupils, that lasts about 15 minutes, in which different opinions are expressed, and the functioning of the robot is discussed. Below we highlight some interesting passages.

4.2.3. Step 2 - The Drunk Bot is like a special elevator

To clarify her position, and convince her pals, C1 presents an interesting example:

112. C1: The Drunk Bot is like a sort of elevator where there are 100 buttons, but we do not know to which floor each button corresponds [...] and we just push a random button.

114. C1: for me it is random, because...one button is like any other, but it is not random for the elevator because it knows which floor to go.

C1 is comparing the Drunk Bot with a special elevator, with no inscription on the buttons; such an elevator moves randomly from the point of view of a user, but from its point of view it does not move random. However, such an explanation is not enough to convince C1’s friends, and the discussion goes on.

4.2.4. Step 3 - Using different random generators

We observe that C1 associates a random phenomenon, related to the Drunk bot, to another random phenomenon, related to an elevator, showing an ability to connect and compare different random generators. This we believe to be a positive result, because literature on the subject had shown that pupils may have difficulties in interpreting different random phenomena as all representing randomness. Rather they may tend to interpret them as totally disconnected phenomena.

We found some more data on this issue. In fact one of the pupils recalls a special situation in which the class substituted, with a coin, the special random generator of the Drunk Bot. The movements of the Robot were still the same then, but the direction to be taken was chosen by means of throwing a coin, instead of using the Roller system of the robot, which depends on the movements of a marble.

136. C2: what about when we used the coin?

137. C1: it [Drunk Bot] moved randomly!

This excerpt witnesses again the pupils ability to make connections between different

random phenomena, moreover it suggests to us that the study of a unique random phenomenon (the movements of the Drunk Bot), which is driven by different random generators (either the coin or the Roller, or other system) can help pupils to interpret different random generators under the same idea of randomness. In other words, we start from different random generators, and we use them as interchangeable parts of a unique random phenomenon; this provides pupils with a natural connection between the different random generators.

4.2.5. Step 4 - The Drunk Bot is a mixed thing

The discussion started by C1 ends up with a pupil, C3 clarifying C1's ideas:

166. C3: [...] C1 means to say that [...] where the ball goes is random, while the movement done by the robot is not random, but however it is dictated by the movement of the ball, which is random

167. C3: it is a random thing that we move non randomly

168. C4: it is a mixed thing

In other words pupils are able to distinguish which element of the Drunk Bot are random and which are not; they are able to decompose the phenomenon into a random part and into a non random part, which we again consider to be a meaningful result in terms of the ability to individuate randomness in given phenomena.

5. Conclusions

The analysed data suggests that the ability of changing standpoints and also taking external standpoints, can give insights into the complexity of random phenomena. In particular it may allow the pupil to individuate the random and non random components of a complex phenomena on the one hand, and on the other hand to compare different phenomena by comparing their random components. We believe that the attitude, and capability, to consider different standpoints, can be (as in our case) fostered by proposing pupils activities involving physical microworlds, which are external from the pupil, allowing a detachment from the phenomenon.

The second key indication we abstracted from the data is derived by observing that pupils actually individuated the random generator of the Drunk Bot, and hypothetically substituted it with another random generator. Such substitution was functional to the ongoing class discussion aimed at classifying the drunk bot in terms of being random or non random. The pupils conclude the discussion agreeing on considering the robot as a mixed entity, both random and non-random. In this passage we believe that a key role was played on the one hand by the request of classifying the robot, and on the other hand by the design rational underlying the random phenomena proposed in the activity sequence. In fact each proposed phenomenon has a random generator which some how dictates the behaviour of the other parts which are not actually random, as clearly explained by C1 in the reported class discussion. In this perspective, the random generator of a phenomenon, can be "taken out" and substituted with another random generator, taken from another phenomenon, as in the case of the coin used to "drive" the Drunk Bot. If that is the case, we argue that the fact that the two different random generators are employed as equivalent random components dictating a complex phenomenon, may foster the building of connections

between the meanings raising from the study of each of the two random generators. We plan to test this hypothesis in the rest of our experimentstion which will be based on computer microworlds that will be designed ad hoc following the principles we presented in this paper.

References

- Bartolini Bussi M. G.(1996). Mathematical Discussion and Perspective Drawing in Primary School. In *Educational Studies in Mathematics*, 31 (1-2), 11-41.
- Bottino, R. M.: 2001, *Advanced Learning Environments: Changed Vies and Future Perspectives*. In "Computers and Education Towards an Interconnected Society", pp. 11-26. Edited by M. Ortega, J. Bravo. Kluwer Academic Publishers, Dordrecht/Boston/London.
- Cerulli, M., Mariotti, M. A. (2003): *Building theories: working in a microworld and writing the mathematical notebook*. In "Proceedings of the 2003 Joint Meeting of PME and PMENA". Vol. II, pp. 181-188. Edited by Neil A. Pateman, Barbara J. Dougherty, Joseph Zilliox. CRDG, College of Education, University of Hawai'i, Honolulu, HI, USA.
- Fischbein, E. (1983). Intuition and Proof. *For the Learning of Mathematics*, 3(2), 9-19.
- Fischbein, E. (1975). *The Intuitive Sources of Probabilistic Thinking in Children*: Reidel.
- Kahn, K. (in press). "ToonTalk - Steps Towards Ideal Computer-Based Learning Environments". In *A Learning Zone of One's Own: Sharing Representations and Flow in Collaborative Learning Environments*. Mario Tokoro and Luc Steels, editors, Ios Pr Inc, June 2004.
- Konold, C. (1995). Confessions of a Coin Flipper and Would-Be Instructor. *The American Statistician*, 49(2), 203-209.
- 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". *Mathematics Education Library*, vol. 17. Kluwer Academic Publishers, Dordrecht/Boston/London.
- Papert, S. (2000): *What's the big idea? Toward a pedagogy of idea power*. IBM Systems Journal, Vol 39, NOS 3&4, 2000.
- Pratt, D. (1998) *The Construction of Meanings IN and FOR a Stochastic Domain of Abstraction*, Unpublished Ph.D. thesis, 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. Lawrence Erlbaum Associates, Inc.
- Truran, J. M. (2001). *The teaching and Learning of Probability, with Special Reference to South Australian Schools from 1959-1994*. Unpublished Ph.D. thesis, Faculty of Arts – Graduate School of Education, Faculty of Mathematical Sciences – Department of Pure Mathematics. University of Adelaide.
- Volchan, B. S. (2002): *What is a Random Sequence?.* The Mathematical Association of America, Monthly 109, January 2002, pg. 46-63.
- Wilensky, U. (1993). *Connected Mathematics - Building Concrete Relationships with Mathematical Knowledge*. Unpublished PhD Thesis, Massachusetts Institute of Technology.
- Wilensky, U. (1997). "What is Normal Anyway? Therapy for Epistemological Anxiety". *Educational Studies in Mathematics*, Special Issue on Computational Environments in Mathematics Education, Volume 33, No. 2. pp. 171-202. Noss R. (Ed.).

STUDENTS' MEANING-MAKING PROCESSES OF RANDOM PHENOMENA IN AN ICT-ENVIRONMENT

Kjærand Iversen, *Agder University College, Norway*

Per Nilsson, *Växjö University, Sweden*

Abstract: *This paper brings to a focus the different ways in which lower secondary students handle compound stochastic phenomena. The analysis is based on clinical interviews in which the participants explore different several-step problems in an ICT-environment. How the students understand the content within the situation is regarded from the perspective of how their understanding varies with their interpretation of the situation. A leakage strategy and a division strategy are identified as being of particular importance for several students in their meaning-making processes.*

Background

Peoples' different ways of handling chance encounters has been the object of a great number of studies. Shaughnessy (1992) and Gilovich *et al.* (2002), among others, provide overviews of the field. Gilovich *et al.* focus mainly on the psychological approach towards heuristics and biases, whereas Shaughnessy also discusses some results linked to educational issues. Despite the extensive research on pupils' encounters with probabilistic reasoning we claim that there is need for further investigation. In particular we wish to discuss issues of students' dealing with probabilistic phenomena in an exploratory and interactive setting.

The current paper focuses on students' reasoning processes with respect to *compound stochastic encounters*, brought to the fore in a computer-based environment. More precisely, our concern is on how students, in an ICT-environment, explore and make use of different strategies when handling random processes, which can be divided into several steps.

In the following part we will define and explore, in a more precise way, the stochastic phenomenon, which is of interest to us.

Simple and compound stochastic phenomena

Stochastic phenomena involve one or more *stochastic objects* (random generators). In contrast to deterministic phenomena, where there is only one outcome, several different outcomes are possible in a stochastic phenomenon.

Stochastic objects are often put into action by an external force or mechanism, but it is the objects' inherent characteristics, such as form, symmetry, centre of gravity and so on that determine the probabilities for the possible outcomes. A *simple* stochastic phenomenon involves only one object and it is put into action only once. Throwing one die exemplifies this. If the phenomenon is not simple it is called compound. A

simple series of throwing a die illustrates this concept. However, this paper focuses on a particular kind of compound stochastic phenomena where one object moves continuously from a start position to and end position, encountering one or several bifurcations along the paths. These phenomena we name *one-object stochastic phenomena* (OOSP). The robot problem used by Green (1983) illustrates this.

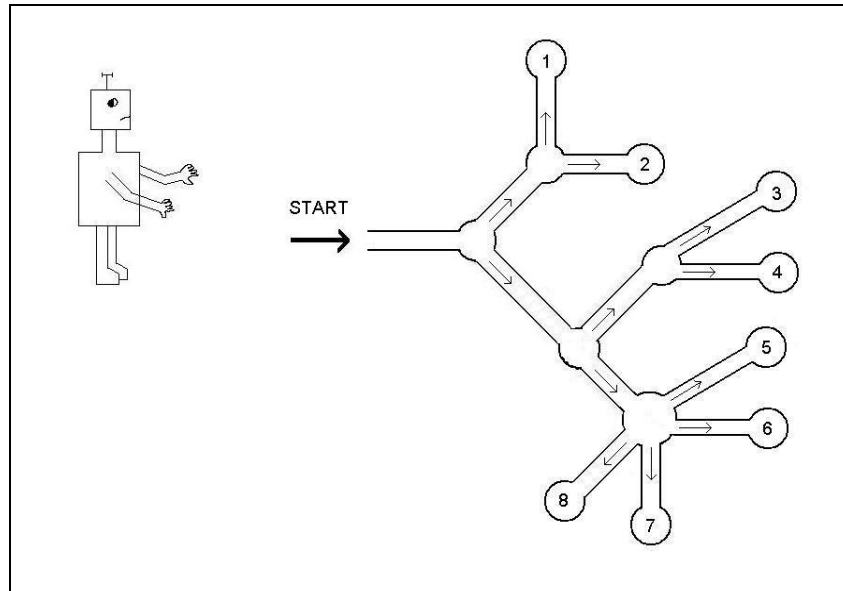


Figure 1. The maze in the Robot-problem in Green's survey.

In this situation a robot is walking into a maze. Sometimes the robot encounters a crossroad and the continuation is decided by a random mechanism. Finally the robot ends up in one of eight rooms. In Green's survey the following question was asked: "In which room is the robot most likely to finish up, or are all rooms equally alike?"

From an expert's point of view this problem might be resolved by using the *Product Law*: If two events A and B are given then: $P(A \cap B) = P(A) \cdot P(B | A)$. Using this law, a compound problem may be split up into sub-steps after which individual conditional probabilities may be multiplied. Stochastic situations where this procedure may be used are commonly called *several-step problems*.

In an attempt to establish the issue of the current study, regarding students' ways of encountering OOSP, we first present two prominent results from earlier research.

Previous research

As reported by Green (1983), the result in the robot problem was surprisingly poor for all participating students (age 11-16). Only 13% of the grade 10 students (age 15-16) gave a correct response, even though these students had been working with tree diagrams and the product law of probability.

Considering chance encounters, Fischbein (1975) emphasizes the role of intuitive reasoning. He distinguishes between primary intuitions as cognitive acquisitions, derived from individual experiences, without systematical instruction, and secondary

intuitions as formed by education and linked to formal knowledge. According to Fischbein, developing secondary intuitions can be seen as a process in which the student learns to make use of generative mental models. He exemplifies this line of reasoning with the structural potential in using the well-known tree diagram. In this connection, Fischbein *et al.* (1975) were using six different devices to explore students' probabilistic intuitions. Their results were quite contrary to the results of Green's study (as pointed out by Green). In an experimental environment (with manipulative materials) students were given problems similar to the robot problem. A majority of the students gave a correct response to most OOSP problems. In the last (see Figure 2b), it was only among the oldest students that a majority (more than 50% of the students) gave a correct response to the question.

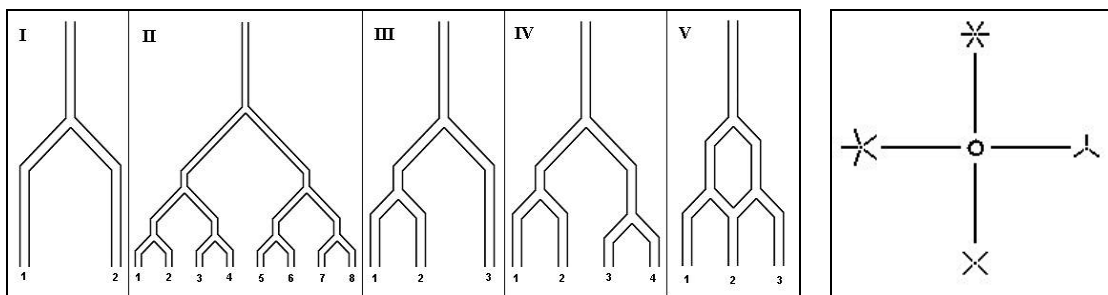


Figure 2. Random devices used by Fischbein

a) The five "two-dimensional" devices. When a marble is released from the top it passes through one or more in the first; 3, 4, 5 or 6 in the second crossroads on its way down.

b) Schema of a three-dimensional device. The marble has four options in the first; 3, 4, 5 or 6 in the second crossroad.

The study by Fischbein *et al.* (1975) included pre-school children (age 5-7), and surprisingly two of the tasks (II and IV) were solved better by them than by the oldest students (age 13-14). The authors explain this as a result of education, claiming that schoolwork often seems to orient the child towards deterministic interpretations of phenomena.

In the current study, correct or incorrect responses are not of particular interest. Our interest is directed towards a more qualitative analysis of students' ways of dealing with OOSP, in order to better understand different kinds of responses given and strategies used. In particular, we focus on how students interpret the tasks, which problems they engage themselves in, and relate this to their articulated explanations and strategies.

Theoretical considerations

A long time ago Bruner (1968, p. 4) stated,

“...when children give wrong answers it is not so often that they are wrong as they are answering another question...”

The meaning of this statement is that, if we are going to understand students' ways of encountering random phenomena we have to take into account the questions the

students are engaged in. If we are interested in learning objects we have to treat the learners individually and, from time to time, try to ascertain what each individual is trying to accomplish. For this purpose, Halldén (1988, p. 125) saw it fruitful to introduce a distinction between *task* and *problem*. He defines task as:

“what is presented to the pupils by the teacher with the intention that they are to do something and/or that they are to learn something”,

and problem as the learner’s personal interpretation of the task given. Viewing the activity from a normative perspective may restrict the analysis too much, and obstacles rather than possibilities might be considered. Instead, we consider it more fruitful to base the analysis on students’ interpretation of the situation at hand, in order to illuminate students’ strategies and learning potential.

Adopting such a student oriented perspective, implies a necessity to reflect on contextual influences on learning. In accordance with a constructivist view, by context and contextual elements we refer to students’ personal constructions. That is, we consider context as a mental device, shaped by individual interpretations. If we let the *conceptual* context denote personal constructions of concepts, embedded in a study situation, and let the *situational* context denote interpretations of the setting in which learning occurs, and let the *cultural* context refer to constructions of discursive rules and patterns of behavior, we can talk of students’ ways of handling a learning situation as a problem of contextualization (Halldén, 1999; Nilsson, 2004).

Halldén (1999) stresses that these different kinds of contexts are in play simultaneously as we are trying to solve a task but, depending on how we interpret the situation, by focusing on certain aspects, they get different priorities in the contextualization process.

Considering situational contextualizations specifically, one could argue that an ICT environment offers interesting possibilities. It is an arena where the concrete and the abstract, or the informal and formal, can be related. This gives ample opportunities for the students to integrate pieces or fragments of knowledge. Pratt (1999, p. 61) writes:

“Computers provide a medium for designing activities that build and integrate pieces of knowledge. A microworld may be able to integrate these fragments of knowledge by offering opportunities for their use, enabling the construction of meaning.”

Difficulties in handling tasks presented in a study situation, and in acquiring new concepts and strategies, are thus seen as students’ difficulties in contextualizing the task in such a way that new information can be interpreted and new strategies worked out. The learners’ ability to discern which relevant parameters in a learning situation will be of crucial importance and also, their ability to evaluate different ways of viewing the world, that is, their abilities to differentiate between different contextualizations of a given task. Caravita and Halldén (1994, p. 106) express it the following way:

“Learning is then a process of decentering, in the Piagetian sense, rather than the acquisition of more embracing logical or conceptual systems replacing earlier less potent ones.”

Due to limitations, we are not presenting any details regarding processes of decentering. Instead, we focus on strategies that appear in a computer-based environment, and how these strategies are related to the students’ different ways of posing the problem, i.e. their interpretation of the study situation and the learning material.

Object of study

The aim of the study is to describe students’ different ways of acting within a specific computer based environment called Flexitree. In particular we are interested in how students’ strategies appear when they explore OOSP-tasks.

Flexitree

The role of the Flexitree-environment is intended to be two-folded. First it just has the role of replacing a real device that would be difficult to build (e.g. Fischbein’s devices). Second it has the power to generate a lot of data in short time, which can be helpful in a frequentist approach to probability.

In the software, the user may choose between several different *Flexitrees* – setups in which marbles can move. At the start the marbles are at the top of the system. When pushing RELEASE the marbles move downwards. They may be temporarily stopped by pushing STOP and then released again by pushing CONTINUE. Finally they end up in one of several boxes at the bottom. A table and a diagram keep track of the number of marbles in each box. The implicit probability model in each crossroad can

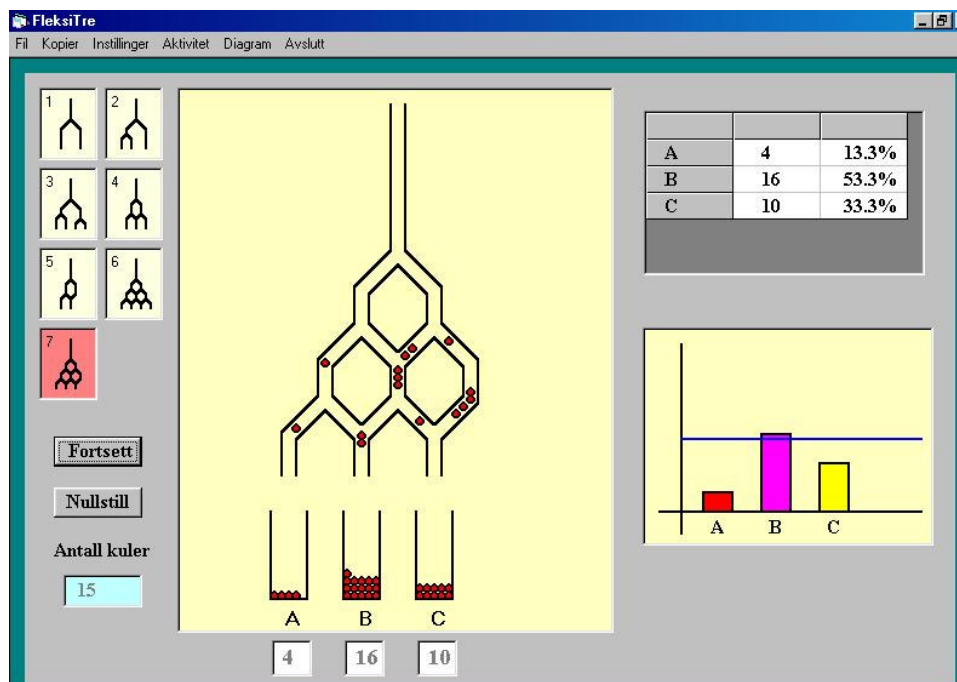


Figure 3. Screenshot with some "action" from Flexitree 7. The other setups are shown as icons to the left.

be determined by the students. However, in this paper we only focus on setups in which the probability is one half for each path at each crossroad.

Method

Semi-structured interviews were used to collect the data. To our view, in agreement with Ginsburg (1981), standard tests or naturalistic observation would not give the in-depth data needed to fulfill the aim of the study. Naturalistic observation seems not to be practical as the waiting time for useful spontaneous verbalizations might be quite long. Also, a written test is not suitable, Ginsburg (1981, p. 7) writes:

“When the underlying cognitive processes are numerous and complex, standard tests may be ineffective or at least inefficient.... A small set of standard questions....will not suffice to capture the richness and complexity of the relevant cognitive structure.”

The participants were students from lower secondary schools (age 14-16). They had some experience with several-step phenomena (but not much). The problems used, were expected to be quite challenging for the students.

The students were sitting in front of a PC. The session started by the interviewer explaining the basic features of the software. Then the students were given some time to become familiar with the features of Flexitree. The implicit probability model was not explained to them.

The screen-image was videotaped along with the sound by using a Televiever. These data were then transferred to a PC. Both the transcript and the digitalized videos were used in the analyses.

Data was collected in several iterations. Based on methodological and theoretical considerations, the setting and protocol for the interview were modified from step to step. This paper concerns the interviews from the first two iterations in the study with 4 respectively 12 students. In iteration one, the students were individually interviewed. This setting seemed to make the students less willing to talk and they were also strongly influenced by the interviewer's suggestions. Hence students were interviewed in pairs in iteration two.

In each setup, the students were asked to focus on two tasks:

T₁: Are the probabilities for ending up in the different boxes equal or not?

T₂: What is the probability for ending up in one particular box?

The students were asked to explain their answers. If the students seemed stuck, or were not making any progress, the researcher intervened with questions or suggestions.

Results

From the analysis of students' exploration with Flexitree, several different strategies were identified. In this paper, we focus on two strategies where the students put quite different meaning to the crossroads. We have chosen to call them the *leakage*

strategy and the *division strategy*. In the following, we will explain those strategies further using two interviews for illustration.

The leakage strategy

Before entering setup 7, Oskar (grade 9, age 14) has tried setups 1, 2 and 4 (see Figure 3). In his previous activities most of his work has been based on guesses, which he has had difficulties to explain. However, taking into consideration frequency data in setup 1, he accepted that half of the balls go left and half of the balls go right in the crossroad?

Setup 7 begins by asking Oskar to come up with a first hypothesis. His guess is that about 50% of the marbles will end up in B and 25% in A and C respectively. This response could be explained by reference to his previous contextualizations of setup 4 and the frequencies encountered while working with this setup.

Oskar then starts working with Flexitree. Observing some relative frequencies, he almost immediately starts to doubt his first guess. Continuing working with Flexitree for a while, he suggests that the probabilities for ending up in boxes A, B and C are $1/6$, $2/6$ and $3/6$ respectively. When the observer asks Oskar for an explanation, Oskar has difficulties in answering. The interviewer then tries to make Oskar focus on a sub-task, namely to explain the low amount of marbles, ending up in box A. In doing so, what we call *the leakage strategy* appears (R stands for researcher):

O: Because when they go ... because the probability for /marbles/ coming to A is quite small ... because ... on their way ... there is only one path to A. And then ... on this path there are two paths that go like this [pointing to the two paths leading off the path to A].

R: Mm.

O: And then it becomes less likely.

R: Yes.

O: And then ... the marbles are disappearing so to speak.

R: Yes.

O: So I can ...

R: Is it (the probability) lower or higher than one sixth?

O: Maybe it's one sixth. Now, at least, I think I understand why the chance is so small for A.

R: Mm. There are fewer and fewer along that path?

O: Yes. There are holes in the path in some sense.

R: Yes.

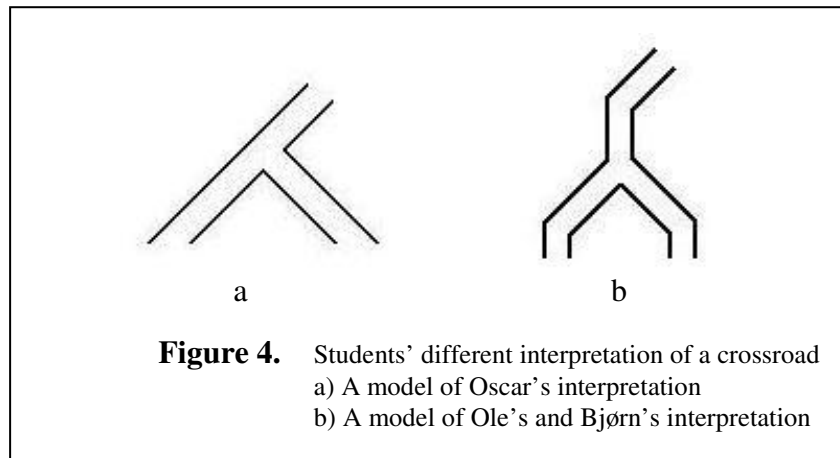
O: And then the marbles ... they fall off the steep slope and then go to B and C instead.

R: Yes. So, on the path downward they become fewer because some are dropping off?

O: Yes. Only a few continue downward.

After the intervention Oskar is not trying to give an explanation to the probability estimate $1/6$. Instead, our interpretation is that he considers the situation as a problem of explaining why so few marbles end up in box A. He tries to do so by focusing on what meaning the crossroads have for the marbles moving toward box A. However, what could be argued is that he seems to interpret these passages as leaking holes

rather than ordinary crossroads. If we, based on this, interpret Oskar's idea of the rolling marbles as similar to what is described in Figure 4a, we understand the



difficulties Oskar has in giving meaning to his probabilistic estimates. Even if he has noticed the importance of the crossroads, in this case his particular contextualization does not allow him to give a numerical meaning to them.

The division strategy

To illustrate the division strategy we will look at a pair of students from iteration 2. After playing a while with setup 2, Ole and Bjørn (grade 10, age 15) are asked to make a first suggestion about the probability for marbles ending up in B. Their response also included an explanation of their hypothesis.

- O: It is ... actually ... as much as ... I believe that there is ... there's a fifty percent chance for marbles coming there [pointing at C].
- B: Yes, it's fifty -fifty [pointing to the first crossing]
- O: Yes, but it is also ... it is ... then it becomes almost a twenty-five percent chance for marbles coming to B and a fifty percent chance for them coming to C.
- B: It must be twenty-five.
- O: I was about to say that there is because fifty is divided by two again. The chance is smaller for them coming to A and B, than there is for them coming to C.
- R: Ok. What is the chance for them coming to B?
- O: I believe it's twenty-five.

The students are aware of the fact that the probability is reduced in a crossroad and since two paths lead from a crossroad they choose to divide by two. The division strategy seems to be dynamic in the sense that the students follow the marbles from the top and towards the end, beginning with 100 percent and then using repeated division by two an appropriate number of times. (In the case above: 100% → 50% → 25%). Their way of reasoning seems to involve aspects of idealization. Based on symmetrical features, the students view the distribution of marbles from a crossroad to be completely equal, that is, they model the crossroads as being a fifty-fifty chance, as shown in Figure 4b.

Discussion

The aim of this paper has been to investigate strategies that become evident in students' different ways of handling OOSP-tasks in an exploratory setting. From the variation in outcomes, it can be argued that the participants interpret the situation in different ways. The participants activate different contexts for interpretation in their meaning-making processes. Two main strategies are brought to the fore. The first has been described in terms of "holes in the way" and we called this approach *the leakage strategy*. The second strategy describes a situation in which the students interpret the task from a computational point of view, in terms of the mathematical operation of division. We have referred to this strategy as *the division strategy*. In our data this strategy was more frequently used by the students than the leakage strategy.

When students are using the leakage strategy, aspects of random phenomena are taken into account in the reasoning process. The students seem to be aware of the fact that the probability is in some sense distributed, and that the marbles are distributed in a non-deterministic manner. However, it seems to be unclear to the students precisely how such phenomena work. As argued above, this seems to be due to the students being involved in a physical interpretation of the situation, similar to what we have modelled in Figure 4a. Compared to the robot problem, illustrated in Figure 1, it would be interesting to further explore the leakage strategy. Based on our result it could be argued that proper answers would be more numerous, if the ways out to room 1-4 had been directed downwards instead of upwards. In such a representation of the problem, the crossroads may be interpreted as leaking holes and by that lowering the chance for marbles to end up in rooms 5-8.

With the division strategy, the students saw a crossroad as a fifty-fifty percent chance. It could be argued that the students, in this case, in a direct way take into consideration the number of paths leaving a crossroad, that is, that they should divide by two. To gain even more information of such reasoning, it would be interesting to create a Flexitree situation, containing features similar to those of Fischbein's pagoda, in which marbles encounter crossroads having three or more exits.

Our results point to some important aspects of the teaching-learning environment. Firstly, teachers have to realize that a learning situation is not static. All students do not perceive the learning object in the way the teacher intends. Secondly, an analysis of students' interpretations in general and in their choice of strategies in particular, may have a pedagogical potential in making teachers aware of the many possible routes that students' activities may take, which, in turn, may improve their capability to intervene in such situations.

References

- Bruner, J. S.: 1968, *Towards a Theory of Instruction*, New York: Norton.
- Caravita, S. & Halldén, O.: 1994, Re-Framing the Problem of Conceptual Change. *Learning and Instruction*, 4, p. 89-111.
- Fischbein, E.: 1975, *The intuitive source of probabilistic thinking in children*, Dordrecht,: Reidel.
- Gilovich, T., Griffin, D., & Kahneman, D. (eds.): 2002, *Heuristics and Biases: The Psychology of Intuitive Judgement*, Cambridge: Cambridge University Press.
- Ginsburg, H.: 1981, The clinical interview in psychological research on mathematical thinking: Aims, rationales, techniques, *For the Learning of Mathematics*, 1(3), p. 4-11.
- Green, D. R.: 1983, A survey of probability concepts in 3000 pupils aged 11-16 years. In D.R. Grey. P. Holmes, V. Barnett & G. M. Constable (Eds.), *Proc. First Intern. Conf. on Teaching Statistics*, p. 766-783, Sheffield, UK: Teaching Statistics Trust.
- Halldén, O.: 1988, Alternative Frameworks and the Concept of Task. Cognitive Constraints in Pupils' Interpretations of Teachers' Assignments, *Scandinavian Journal of Educational Research*, 32, p. 123-140.
- Halldén, O.: 1999, Conceptual Change and Contextualisation. In W. Schnotz, M. Carretero & S. Vosniadou (Eds.), *New perspectives on conceptual change*, p. 53-65, London: Elsevier.
- Nilsson, P.: 2004, Students' ways of conceptualizing aspects of chance embedded in a dice game. In M. J Høines & A. B. Fuglestad (Eds.), *Proc. 28th Conf. of the Intern. Group for the Psychology of Mathematics Education Vol. 3*, p. 425-432, Bergen, Norway.
- Pratt, D.: 1999, *The construction of meaning in and for a stochastic domain of abstraction*, unpublished doctoral dissertation, London: The University of London
- Shaughnessy, M.: 1992, Research in probability and statistics: Reflections and directions. In Grouws, D. A. (eds.), *Handbook of Research on Mathematics Teaching and Learning*, p. 465-494, New York: Macmillan,

YOUNG CHILDREN'S EXPRESSIONS FOR THE LAW OF LARGE NUMBERS

Efi Paparistodemou, *University of Cyprus, Cyprus*

Abstract: *The aim of this research is the design of a game to afford expressive power to children aged six to eight in the domain of probability. Particularly, this paper focuses on how young children express ideas for achieving a fair result by using a computer game. The computer game offered children the opportunity to make their own constructions of sample space and distribution. The children used spontaneously five distinct strategies to express the idea that their construction could only be judged with respect to a large number of trials. It is apparent that the game provided children the opportunity to express the idea that stability can come from increasing outcomes with different strategies. It can be said that young children's expressions is evidence of several 'situated abstractions' for the law of large numbers.*

Introduction

Tversky and Kahneman (1983) coined the phrase 'the law of *small* numbers' to characterize a commonly held yet erroneous belief that the properties of an unknown sample space can be directly assessed through a relatively small number of observations of its constituent elements. In her research on understanding the idea of randomness of primary grade children, kindergartners, 3rd graders and undergraduates, Metz (1998) claims that the law of small numbers typified these subjects' strategies for predicting results of a random behaviour. Focusing on the issues of agency and control in probabilistic situations, she demonstrated that students assumed that they had more agency and control than they in fact had, and that their sampling strategies were informed by these incorrect assumptions. She mentioned that in this strategy the participant believes he or she should somehow be able to implement the drawing process. For example, the order of marbles drawn accurately from a box reflects the proportion in colours in the unknown sample space.

The above research falls broadly into the 'misconceptions' paradigm. In the field of understanding probability, misconceptions were defined by Tversky and Kahneman (1983) as failure to behave rationally. Tversky and Kahneman (1983) suggested that misconceptions stem from prior learning, either in the classroom or from their interaction in the physical and social world. Smith, diSessa and Rochelle (1993) proposed an alternative perspective that refutes the 'misconception approach' on empirical and methodological grounds: they argue that misconceptions are in fact context specific and not general. Moreover, they claim that misconceptions were elicited from people who were being asked to respond to questions outside their areas of competence and/or in the absence of appropriate tools to explore the questions. Smith, diSessa and Rochelle (1993) suggest that instead of searching for

misconceptions, researchers could possibly locate the principled knowledge underlying students' responses and leverage this knowledge in suitable learning environments. Biehler (1991) argues for the potential role of the computer in supporting students' understanding of central constructs in the domain of probability and statistics, for example in dealing with the *law of large numbers* (hence, L.L.N.). According to his view, computer support enables students to transcend their default 'law of *small numbers*' heuristic and gain insight into properties of distributions. Pratt (2000) demonstrated the potential of computers for learning probability with understanding. His 10-and-11 year-old participants manipulated stochastic gadgets, representing everyday objects such as a die, a coin, a lottery machine, and a set of playing cards. Individual learners expressed their beliefs in symbolic (programming) form and articulated their beliefs, construct, and reconstruct probabilistic situations in the light of their experiences. In this study, Pratt (2000) used an approach in which children articulated their meanings for chance through their attempts to 'mend' possibly broken computer-based resources.

My point of departure in the current study is that young students implicitly understand the L.L.N. and given appropriate tools, they will express these intuitions. The aim of the broader research of the present paper is to design such tools and evaluate their efficacy. I adopt Fischbein's (1975) definition of intuitions as *being based on knowledge from experience in order to get control over an action*. In this study I also adopt the notion of *situated abstraction* (Noss, Hoyles, & Pozzi, 2002), defined as a conceptualisation of mathematical knowledge that is simultaneously situated and abstract. The process cannot be separated from the product; and this is part of the webbing idea' (see Noss and Hoyles, 1996).

In this paper, I present 'snapshots' of young children using a computer game and show how the game provided children opportunities to express that the stability of an outcome in random situations increases with the number of trials.

Methodology

A game was designed to afford children the opportunity to talk and think about probability and was built to allow a connection between local and global events¹. The 'lottery machine' (c.f. Pappastodemou & Noss, 2004; see Figure 1) is a visible manipulable engine for the generation of random events. Using this game, children could directly manipulate its outcomes.

¹ A *local* event refers to the trial-by-trial variation and the *global* to the aggregate view of each single trial. Children can use local events to make sense of short-term behaviour of random phenomena, while global events are associated with long-term behaviour. Thus, whereas an individual outcome could be seen as a single trial in a stochastic experiment, the totality of these outcomes gave an aggregated view of the long-term probability of the total events (c.f. Wilensky, 1997 and Dave, 2000).

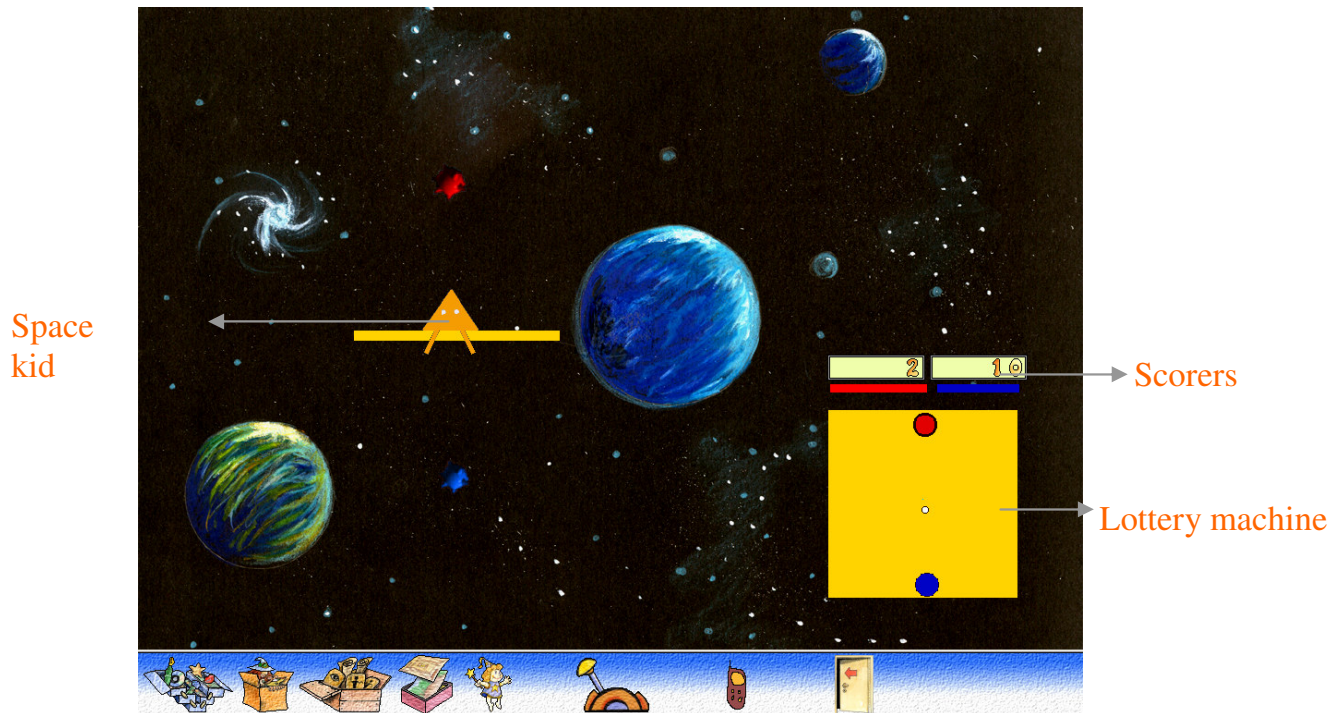


Figure 1. A screenshot of the game.

The children could make changes in the ‘lottery machine’ in which a small white ball bounces and collides continually with a set of blue and red balls. As programmed initially, collisions with the red balls added one point to the red score and moved a ‘space kid’ one step up the screen. In this way the lottery machine controlled the movement of a space kid. In contrast, collisions with the blue balls added one point to the blue score and moved a ‘space kid’ one step downwards. The link between the objects of the game was visible to the children, as the software provided a library of icons in the shape of stones which enabled the user to make rules that would connect one object to another (for more details see Paparistodemou & Noss, 2004). The children could change and manipulate a number of aspects to control the properties of the lottery machine: the number, the size, and/or the position of the balls.

The ‘Space Kid’ game can be seen as a form of ‘random walk’. The random walk in the game involved a ‘space kid’, which moves upwards and downwards on a yellow line. These movements occurred one step at a time, in response to an outcome generated by the lottery machine. The existence of the two planets on the screen (the blue and the red above and below from the yellow line) was for the space kid not to move far away from the yellow line. The planets had the rule ‘when I touch the space kid the game to stop’. As a result, the game stopped each time the space kid reached them (for more details see Paparistodemou, 2004).

The game was designed over three iterative cycles. Twenty-three children, aged between 6 and 8 years, were interviewed during the last iteration. The children interacted with the computer game individually between 4-6 half-hour sessions. The participants had no formal learning experience with probability. In the first task, students were asked to make the space kid move quite close around a centre line in

order to construct a ‘fair sample space’. Table 1 shows an outline of the tasks that

On the screen	Children could manipulate	Goal	Probes	Tasks
Space kid Lottery machine Bouncing ball One blue and one red ball Scorers of the balls Two planets	Balls: number size speed arrangement in lottery machine Space kid: starting point The position of two planets	Keep the space kid near the yellow line.	How do children express their ideas about randomness, fairness? How do children judge their constructions?	What will you change so that the space kid stays near the yellow line? Try it out! What happened? Why do you think this happened? If it doesn't work, as you want it, what else can you change?

were posed during the interviews.

Table 1: An outline of the tasks

Expressions of the law of large numbers

The children expressed in a number of different ways the idea of having a large number of outcomes. The idea of the law of large numbers was expressed when children made judgments whether their own construction in the lottery machine was fair or not. Their expressions can be categorised into: 1. increasing the speed of the white ball, 2. adding more coloured balls in the sample space/distribution, 3. adding more white balls, 4. making the size of the white ball(s) bigger, and 5. leaving the game to work for longer time.

1. Increasing the speed of the white ball

The idea of changing the speed of the white ball occurred when children had made a fair construction in the lottery machine, but could not get fairness in the game. Paul (6 10/12 years) explained:

Paul: Let's see... Oh! We have more blue scores. It moved down. Oh...I will change the speed of the white balls. I won't watch the numbers. It will move too fast!

He takes the star (a tool of the software) and changes the speed of the white balls.

Paul made the white ball move faster than before in order to demonstrate that his construction in the lottery machine worked, since he believed that, in the long term, his sample space was fair.

2. Adding more coloured balls

Getting more points quickly was also expressed by adding more balls inside the sample space. For example, Simon (7 10/12 years) added more balls to his construction (R stands for researcher):

Simon: It moved up now...equal numbers! Oh! Now it moved down.... Let's see if it moves up. You know something. I will add some more balls.

He stops the game.

I will put these balls together...to communicate (*he laughs*).

R: What do you think will happen?

Simon: We are going to have a better result.

Simon as a basic idea had a symmetric fair sample space with the same number of blue and red balls. He finally decided not to change the symmetrical idea of fairness, but to add where there was a blue ball, a red one and where there was a red ball, a blue one. This idea shows that Simon did not want to change the proportion or the structure of the balls in his lottery machine. This idea also shows that he expected that by adding balls to get bigger numbers his idea would work in the long-term. His action to increase the overall number of balls may be indicative of an implicit application of the law of large numbers. Adding more balls was a strategy that children used very often for constructing a fair sample space. This strategy was also very often combined with other strategies in which children generally expressed the idea of having more trials.

3. Adding more white balls

Adding more white balls was a strategy used by nineteen out of twenty-three children in order to make their construction work for 'bigger' numbers. Fiona's (7 years) attempt to get bigger scores was to copy more white balls and make them to touch more easily the coloured balls.

Fiona: It still doesn't work! I think I have to make another change to the balls. Another ball.

R: Will it work with this change?

Fiona: Yeah...the white balls move around and touch all these balls. Ok... and another thing (*she copies more white balls*)...that makes it work! Wand...wand...right! Let's try it on. (*see Figure 2*)

She starts the game.

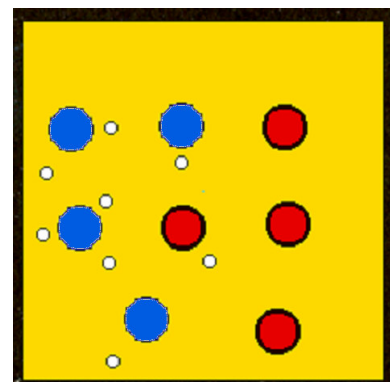


Figure 2. Fiona's construction of fairness to get 'more points'.

Fiona added more coloured balls into her construction and she also added more white balls. Fiona's action can be seen as a *situated abstraction* of the idea that bigger outcomes made her construction more reliable. As she said, her construction would

work better by having more white balls that would make the scorers move more quickly. Fiona appears to recognise that her construction would be good if it would give a fair result in the long-term. She decided neither to change anything in the structure of the coloured balls nor to the proportion between the two colours of balls, but to judge her construction by generating bigger numbers and watching the global outcomes.

4. Making the size of the white ball bigger

Another change to the bouncing ball for getting more points is to change its size. Making the size of the white ball bigger makes the scorers to work more quickly on the screen. This was what Mathew (7 years) did when he wanted to get ‘many points’.

Mathew: ...I will do something else. (*He stops the game*). I will construct two white big balls. I will copy some more red balls, five as the blue ones. (*see Figure 3*).

R: Now, we have 5 reds and 5 blues, how many points will they get?

Mathew: Many points...they will get equal points.

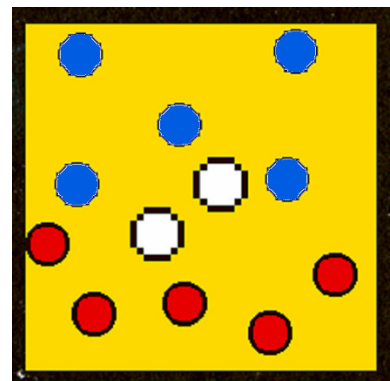


Figure 3. Mathew’s fair construction to get ‘more points’.

Mathew made changes to get bigger number scores by adding more red and blue balls, adding another white ball and making the size of the white balls bigger. Mathew’s statement implies that he is attempting to get equal numbers and there is a need to get large numbers in order to achieve this.

5. Leaving the game running longer

Another idea for getting bigger numbers in the game, is not by making any changes in the mechanism of the game, but by waiting the game to run for a longer time. Orestis (7 years) expressed such an idea about time.

R: How did you arrange them?

Orestis: I mixed them up. Now, they might get equal numbers.
He starts the game.

R: Are they getting equal points?

Orestis: **Not yet.**

Orestis expressed the idea that time is needed to get equal points. His words ‘not yet’ are evidence that he needed time to wait for his construction to work. He implied that his construction must be judged in the long-term and he seemed to believe that time would take care of the (short-term) inequality.

Discussion

It is apparent that the game provided ways in which the children participating in this study could engage with the idea of getting ‘bigger numbers’. The game enabled children to increase the total outcomes by engineering more collisions between the white ball(s) and the red and blue ones. Also students leveraged their intuition for ‘fairness’ as a mental grip on distribution. Thus, the game afforded students *situated abstractions* for the law of large numbers.

In their constructions, children seemed to express a belief that probabilistic ideas, like fairness, e.g. equal likelihood, could only be tested in the long term. This may have resulted from the dual presence of the local events (the colliding balls) and the aggregate outcome (the movement of the space kid). This might add to Metz (1998) finding and it can be said that children by experiencing the game they understood that something that is unstable with a small number of outcomes becomes stable with a large number of trials. The children in this study seemed to express an intuition about stability of long-term trials, a shift of focus that the game promoted by looking at the aggregate outcomes of any construction.

In this paper, I have described young children’s expressions for the idea of the law of large numbers (L.L.N.), which occurred while they built computational models of sample space and distribution. The study demonstrated that students have robust intuitions for the L.L.N., and that – given interactive learning environments – students can express these intuitions. This finding provides support to the general constructionist thesis (Papert, 1991) that engagement in the building of some external, shareable, and personally meaningful product is conducive to mathematics learning.

References

- Biehler, R. (1991). ‘Computers in Probability Education’ in Kapadia, R., and Borovcnik, M. (eds.) *Chance Encounters: Probability in Education* Dordrecht: Kluwer, p.169-211
- Clapham, C. (1990). *The Concise Oxford Dictionary of Mathematics (second edition)* Oxford: Oxford University Press
- Fischbein, E. (1975). *The Intuitive Sources of Probabilistic Thinking in Children* London: Reidel
- Metz, K. (1998). ‘Emergent Understanding and Attribution of Randomness: Comparative Analysis of Reasoning of Primary Grade Children and Undergraduates’ *Cognition and Instruction*, 16, 3, p. 285-365
- Noss, R., & Hoyles, C. (1996). *Windows on Mathematical Meanings: Learning Cultures and Computers*. Dordrecht: Kluwer
- Noss, R., Hoyles, C. and Pozzi, S. (2002). ‘Abstraction in Expertise: A Study of Nurses’ Conceptions of Concentration’ *Journal for Research in Mathematics Education*, 33, 3, p. 204-229

Working Group 5

- Paparistodemou, E. & Noss, R. (2004). 'Designing for Local and Global Meanings of randomness' *Proc.28th Annual Conf. of the Intern. Group for the Psychology of Mathematics Education Vol.3*, Bergen, Norway, p. 497-504
- Paparistodemou, E. (2004). *Children's Expressions of Randomness: Constructing Probabilistic Ideas In an Open Computer Game*, Institute of Education, University of London (PhD Thesis)
- Papert, S. (1991). 'Situating Constructionism' in Harel, I. and Papert, S. *Constructionism* New Jersey: ABLEX, p. 1-11
- Pratt, D. (2000). 'Making sense of the Total of Two Dice' *Journal for Research in Mathematics Education*, 31, 5, p. 602-625
- Smith, J., diSessa, A., and Rochelle, J. (1993). 'Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition' *The Journal of the Learning Sciences*, 3, 2, p. 115-163
- Tversky, A. and Kahneman, D. (1983). 'Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgement' *Psychological Review*, 90, 4, p. 293-315
- Wilensky, U. (1997). 'What is normal anyway? Therapy for epistemological anxiety' *Educational Studies in Mathematics*, 33, p. 171-202

TOWARDS THE DESIGN OF TOOLS FOR THE ORGANIZATION OF THE STOCHASTIC

Dave Pratt, *University of Warwick, United Kingdom*

Theodosia Prodromou, *University of Warwick, United Kingdom*

Abstract: *This paper reports on one aspect of the ongoing doctoral research of the first named author. This study builds on prior work, which identified that students of age 11 years had sound intuitions for short-term randomness but had few tools for articulating patterns in longer-term randomness. This previous work did however identify the construction of new causal meanings for distribution when they interacted with a computer-based microworld. Through a design research methodology, we are building new microworlds that aspire to capture how students might use knowledge about the deterministic to explain probability distribution as an emergent phenomenon. In this paper, we report on some insights gained from early iterations and show how we have embodied these ideas into a new microworld, not yet tested with students.*

1. Emergent phenomena

We begin by considering the relationship between our appreciation of the deterministic, of the stochastic and of emergence and consider how these transitions help or hinder the gradual evolution of the conception¹ of distribution.

Distribution can be seen as a structure with which students can understand all the aggregate features of data sets (Cobb, 1999). Features such as average and spread can be construed as parameters of a theoretical distribution or as emergent from numerous trials. When writers refer to distribution they often shift between these two meanings without making that shift explicit. In this sense, the relationship between these two construals is slippery, and yet perhaps that ambiguity is at the very heart of a deep appreciation of stochastics.

Based on Wilensky's work (1997), we define probability distribution as an emergent phenomenon (Prodromou, 2004). In this way, we hope to capture both the theoretical and emergent perspectives. The distribution would appear as the outcome from many random events and yet there is a sense of organisation represented by the theoretical parameters.

In the service of making sense of the world, people appear to have an intrinsic desire to attribute meaning to what they observe, a search which leads, in turn, to organisation, the formation of patterns, the encoding of pictures, and simplification. Even complex dynamic systems are simplified into emergent phenomena – that is functional collectives which arise through the co-specifying activities of numerous *micro* agents. These sorts of phenomena might be further

¹ By "evolution", we wish to focus on an individual's thinking-in-change (Noss & Hoyles, 1996) as s/he uses the tools that we are designing. Hence we refer to evolution with respect to an individual's thinking and we refer to conception (rather than concept) to emphasise that particular person's construction of the idea.

described as “bottom-up”, as emergent macrobehaviours emerge from behaviours and localised rules of individual agents. Discussions of emergent phenomena are often accompanied by such classic examples as the flocking of birds.

In more technical language, emergent behaviour is evident within a complex adaptive system, in which for reasons that are not fully understood, low level *agents* produce behaviour of a higher sophistication. Such phenomena are not deterministic, do not readily submit to analytic methods and cannot be strictly understood through means of analysis

Emergent phenomena and artificial emergence open up a novel perspective on our interconnected world, indicating that they dominate our life, and will drive the fundamental questions that form our view of the world in the coming era.

The study of emergent phenomena as complex systems is not just considered as a broad new field of science, but as a new framework, a dynamic, revolutionary way of experiencing scientific content. As Davis, B. and Simmt, E. (2003) note, “a different attitude is required for their study, one that makes it possible to attend to their ever-shifting characters and that enables researchers to regard such systems, all at once, as coherent unities, as collections of coherent unities, and (likely) as agents within grander unities” (p.140). An understanding of complex systems is increasingly becoming a core part of scientific knowledge and the adoption of this new perspective is essential to comprehend the world.

We aim to find out how students understand distribution in a rule-governed system. As Johnson (2001) writes, emergent behaviours, like games, are all about living within the boundaries defined by rules, but also using that space to create something greater than the sums of its parts. Indeed, our research aim is to design a microworld setting in which we can observe students harnessing their deterministic causally-based thinking to imbue meaning to probability distribution as an emergent phenomenon. In this respect, distribution throws up some particular pedagogic challenges. The first challenge is to help students to construct through distribution an organising conceptual structure for thinking about variability located within a more general context of data sets (Petrosino et. al., 2003). The second challenge is to help students to discriminate and move smoothly between data as a series of random outcomes at the micro level and the shape of distribution as an emergent phenomenon at the macro level.

Based on Wilensky’s work (1997) about connected mathematics and emergent phenomena, we see probability distribution, both discrete and continuous, as a dynamic and complex construct with a coherent personality (Prodromou, 2004). This personality self organizes out of many individual decisions (data) and a global order emerges out of uncoordinated local interactions over its duration. A pattern (probability distribution) emerges out of the anarchy of randomness.

Trying to make sense of these emergent behaviours is in fact a challenging task (Resnick, 1991). It is noteworthy that the mind struggles to grasp emergent

phenomena successfully, because it boggles at this mix of stability and randomness. We see resonance between the tendency to adopt a centralised mindset when interpreting emergent phenomena and Piaget's seminal work (1975, translated from original in 1951) that reported how the organism fails in the first place to apply operational thinking to the task of constructing meaning for random phenomena. Only much later do we, according to Piaget, operationalise randomness through the invention of probability. In this sense, Piaget offers us a first hint that we only begin to gain some mastery over the stochastic when we learn how to exploit our well-established appreciation of the deterministic. Pratt and Noss (2002) reported that students age 11 years articulated meanings for short-term randomness that were pretty well consistent with those of experts. They were able to discuss randomness in terms of unpredictability, lack of control and fairness in much the same way as statistically aware adults might. Nevertheless, they had little or no language for discussing distribution or the *Law of Large Numbers*. In other words, their appreciation of the patterns that emerge in the longer term was not well developed. As his students worked with a domain of stochastic abstraction, Chance-Maker, they began to articulate new meanings for the longer term. These meanings were causally-based and situated. For example, the students would explain "the more trials, the more even the pie chart". The study described in this paper builds on those ideas and as it attempts to clarify how students at one level let go of the deterministic whilst at the same time re-apply such ideas in new ways to construct probability distribution as an emergent phenomenon.

Much depends of course on the design of the microworld, which must somehow offer us as researchers a *window* (Noss and Hoyles, 1996) on the evolution of such thinking. The microworlds must capture the student's thinking process, or at least a meaningful element of it, by providing sufficient perturbation that we can observe as *thinking-in-change*.

Our broad aim is to observe thinking about emergent distribution in relation to emergent tools. In this particular paper we try to illustrate how insights into thinking-in-change about distribution have informed one design element.

2. Method

The approach of the current ongoing study falls into the category of *design experiments* (Cobb et al, 2003), a methodology that is sensitising us to the complex *learning ecology* through iterative design of the microworld. As we work with students using successive iterations of the software, we are beginning to recognise the emergence of patterns in participants' reasoning evoked by their interactions with the model.

We conjectured that, if participants encounter emergence as letting go from determinism, they might think about distributions as complex adaptive systems and somehow harness meanings for the deterministic in making sense of emergent

behaviour. The challenge was, and remains, one of formulating a microworld design that embodies at each stage testable conjectures, based on our sensitised appreciation of students' meanings as they shift between the micro and macro levels.

We therefore presented the two levels, micro and macro, in separate projects. In the first project, we fore-grounded causality and in the second project we placed emphasis on emergent distribution. The students first experimented with the micro level project but even at that level they were encouraged to let go of determinism through the introduction of error in the determining variables. We have been busy developing a meaningful context in which the students would be able to articulate meanings at these two levels. We developed a basketball context.

2.1 Micro level project

In this paper, we wish to focus on insights gained during use of the macro level project and so we will deal rather superficially with the micro. At the micro level, the student was challenged to throw an on-screen ball into the basket. The student can alter the basket size as well as the speed and direction of the throw.

The task directed attention of the student to causality (speed and angle of throw). Then the student was encouraged to let go of causality, by introducing an error factor in throws. This was a fairly natural step since it felt inappropriate that once a successful throw was discovered, the thrower would succeed every time (the world so far being completely determined). Once error had been introduced, the student was no longer completely in control and aspects of randomness had to be addressed.

2.2 Macro level project

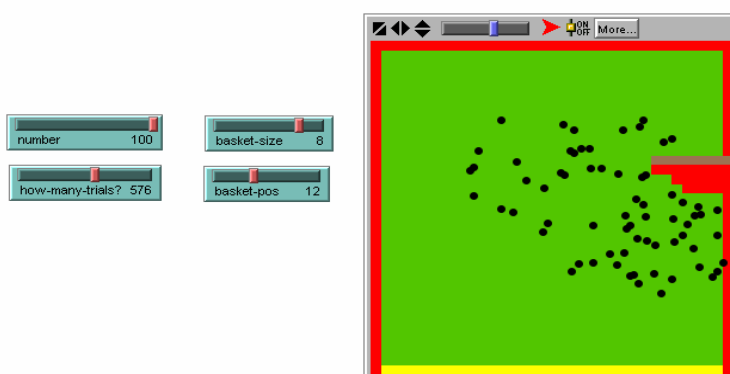
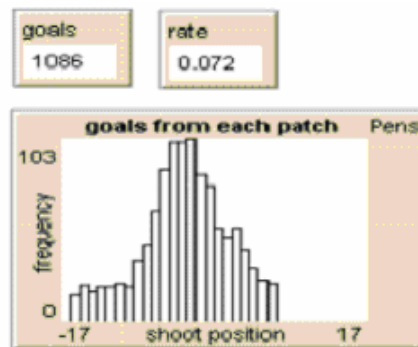


Figure 1: The balls are drawn in black and are travelling in parabolic motion towards the basket. Some are got in and others are failed. The right hand sliders again allow the student to change the size or position of the basket. The left hand sliders enable the student to decide how many balls should be thrown and how many repetitions are carried out.

Whereas at the micro-level the purpose of the task was to succeed in throwing the ball into the basket, at the macro-level the students were asked to design the court and decide how far away the basket should be from the throwers, so a class of unknown children might find scoring neither too easy nor too difficult. The process of throwing was no longer under control of the student but was automated by the computer. This task sought to redirect

attention towards the emergent distribution. We hoped that the student would have

to let go of causality and consider the distribution of the throws. These ideas were incorporated into the macro level project (Figure 1), developed in NetLogo.



At
bot
Figure 2: At macro level the students have access to a histogram of goals from each patch against position of basketball throw.

h micro and macro levels, the students can base their decision from where to throw the balls into the basket, on various types of feedback, such as counters of goals, rate of goals, and three different types of graphs, namely average rate of goals per trial, a histogram of goals against position of basketball throw, and successful-shoots against trials (Figure 2). We focus on the work carried out by two girls, Kate and Anna, (aged between 17 and 18 years) as they engaged with the macro level, having already experienced throwing the ball into the basket at the micro-level. We captured their on-screen activity on video-tape and transcribed those sections to generate plain accounts of the sessions. Subsequently we analysed the transcriptions in attempts to account for the students' actions and articulations. The excerpts in the next section are taken directly from transcriptions of the videotape. ('Res' refers to the first named author.)

3. At the macro level

The two girls chose to work initially with just one screen-child, thus replicating the one thrower situation from the micro-level project. However, they had not anticipated not being able to control the position of that child in the way that they had controlled their own position in the micro-level project. Each time the child threw a ball, it threw from a different place, making it difficult for them to determine success rates in relation to pre-determined positions.

1. Kate: That rate is not very constant... because it is in random places, I think.
2. Anna: We cannot judge because we have only one child... [long pause] ...because he is only a child ... and he doesn't know how to play it.
3. Kate: We have not got certain angle and speed.
4. Anna: We cannot control the throw of the ball.
5. Kate: The rate is very very low so... the line is very down. Almost zero... It is very constant... it is not a straight line.
6. Anna: Ehm...Can we exert any control? There must be a position from where they can score best... Can we find it by standing only on a certain position?
7. Kate: ...but we do not have a fixed shooting position.

The only control that Anna and Kate might have exerted was through the position or size of the basket. The basket however is the end of the process, the result of throwing. Perhaps then they did not construct the basket position as a control. Instead, they were confused by the lack of any control over the position of throws of the unknown child, which were perhaps more readily constructed as inputs or parameters.

This theme of lack of control was widened to encompass the situation as a whole.

8. Res: Ok, so you cannot improve the scoring rate?
9. Anna: Not really.
10. Kate: You could but you can ...randomly, you cannot plan to. It depends on the height they are really and how they throw the balls, and we don't know how they throw the balls.
11. Anna: It is like training different children every time and you don't know how they react because they can really... The only thing you can do is to make the basket size bigger.
12. Kate: We can make the basket bigger or move the basket, but the children don't know that we can do that.
13. Res: Can you improve the performance or control it in a way?
14. Kate: Not really, because there are so many different positions... it is only a small number of people, but the same on different positions... so you don't know where they are going to stand... unless you fix their positions...it's by chance.
15. Kate: It's a bit annoying really, because there are too many variables. Everything is... different every time...the people being standing in different places every time...and they have different heights every time... Everything happens by chance.

The girls struggled to attend to all these variables, which function in parallel and interact independently at a low level. They were unable at this stage to perceive any underlying distribution nor to discriminate and make strong connections between micro and macro levels. Later though they called upon knowledge of bell-curves to begin to recognise features of the distribution.

16. Kate: That's different than before.
17. Anna: That's weird.
18. *Researcher asks her to explain what she means. Anna shows with the cursor on the histogram.*
19. Anna: That's the best position. That graph... the positions towards the middle are better than the ones to the other sides.
20. Kate: Normal distribution.
21. Res: Normal?
22. Kate: Ehm... maybe different because we didn't take too many trials.
23. Anna: And because there are quite... always change... so it cannot be the same every time anyway.

Kate emphasized the macro behaviour and regarded distribution as a coherent unity, but acknowledging at the same time that its behaviour is dependent on the number of trials. On the other hand, Anna linked the macro and micro levels, putting emphasis on the micro level.

At one stage, the researcher asked the girls about the various types of graphs, and we see them interpreting features of the graphs.

Working Group 5

24. Res: What about the three graphs?
25. Anna: They are constant. The graphs are constant, straight lines... (*referring to the successful shots and average rate of goals per trial graphs*).
26. *The researcher asked about the goals for each patch graph.*
27. Anna: That makes sense really... that makes sense because the further away from the basket the harder it is to score, but afterwards... where it goes down are those who scored and the rest would be under the basket or nearly under the basket?

Later, the girls noticed that the distribution can sometimes be skewed.

28. Res: Ok, when do you think we have a skewed distribution?
29. Anna: When you have few trials and few people, and then the distribution is skewed... or few trials really... actually.
30. *They carry out 35 trials with 40 players/shooting positions.*
31. Kate: The distribution is not skewed now, but before it was.
32. Res: You told me that with less trials we will have a skewed distribution, but we don't have. Why?
33. Kate: Maybe that one was by chance or... I do not know.
34. Res: What do you mean "by chance"?
35. *They replicate the same simulation several times.*
36. Anna: Anyway, it always changes...
37. Kate: Yes it is different every time and the rate is much higher than it was the last time... because the height is continually changing... so... I think it is just chance.

Kate and Anna appeared not only to be recognising macro-level features but they were relating those features to how many throws were being used. The notion that "the more trials, the less skewed the distribution" was articulated and seen as something that by chance might not happen. Indeed this pattern was not seen as something that could be controlled.

38. Res: What do you think affects the rate and the shape of the distribution?
39. Anna: I don't think it's controlled to be fair when you've got a bit number of trials and people... because the people are in different places every time... and they have different heights every time... so you can't control it.

Control seemed to be seen in strict deterministic terms, even though they had perhaps begun to articulate earlier a sense of being in partial control of the distribution through the number of throws. It could be argued that this reliance on deterministic control is reinforced through using a computer, and there is some evidence to support that view.

40. Res: You said it will be skewed by chance. Will it be skewed by chance?
41. Anna: (*laughing*) It is probably not by chance, because you have programmed it and you have probably got some complicated formulas... that's why it behaves like that.

However, an alternative argument is that the use of a computer provokes deterministic meanings (cf Pratt, 1998) that we might be able to harness in a productive way in future iterations, much as, according to Piaget, the organism operationalises the stochastic at the stage of formal operations.

4. Key design principles

The episode taken from work with the macro-level project gave us some insights that have allowed us to generate some new conjectures, which we are currently

building into the next design of the emerging microworld. The students continued to see control as embedded in the action of throwing (lines 2-4, 6-7, 10-11, 14, 37) but nevertheless articulated ideas that from our perspective sounded like a form of stochastic control (lines 19-22, 24-27, 29-32). They recognised the bell-curve feature (lines 19-20) and abstracted the notion that “the more trials, the less skewed the distribution”. We would hope though that a more fine-grained appreciation of the nature of control at the macro-level might be constructed. We envisage that a more gradual letting go of deterministic control might allow the students to construct a relationship between the degree of control and the spread of the distribution.

The macro-level project appeared to contain too much randomness, with too little control over how quickly that randomness was introduced (lines 15, 39). We conjecture that the use of two separate models (micro and macro) creates something of an obstacle to shifting between the two levels and so our next design will incorporate the two levels into a single project.

As shown above, the notion of control, or lack of control, was crucial. We therefore plan to embody error in a consistent way across the new single project. Variables such as shooting position, speed and angle will all be fixed by default but with the option of adding error, which itself can be increased or decreased in size. Thus, a student may choose to have only one variable with error and they might choose to gradually increase that error before introducing a second variable with error. We believe that with this additional control over control students might be able to connect the deterministic to the stochastic in a more fine-grained way.

Throughout their use of the micro and macro projects, the discussion about the histograms of goals from each patch seemed to involve the students in beginning to conceptualise probability distribution as an emergent phenomenon. In contrast, discussion about the other graphs was at best trivial and at worst a distraction from our research agenda (lines 25-27). We have realised that some of the graphs are more important in terms of making connections between the levels. We are therefore reducing the number of types of graphs to the most relevant ones but offering those graphs for any or all variables.

To help users focus effectively on critical ideas, it is envisaged that students will be asked to select and look at a graph of any variable that is designed to incorporate an error element. In all of these cases, students can have access to a range of graphs. The default shows a graph of success ratio against time. Students can choose to see a histogram of number of successes against position of scoring, a histogram of frequency of successes against angle, speed or height.

5. A form of stochastic control

In the spirit of design research, we conclude by capturing our current state of understanding in a specification of the new design. In particular, insights into

thinking-in-change about distribution have informed one design element, as follows.

A consistent approach across variables is implemented in a way that all variables are now controlled similarly. We illustrate that approach through Figure 3, which demonstrate the case of release angle and speed. In Figure 3, the release speed has no error and therefore the speed of throw will be entirely determined, never varying from throw to throw. In contrast, the user has chosen to incorporate error into the angle of throw. As a result, two new marks, in the form of arrows, have appeared either side of the slider button. These arrows can be moved independently of each other and of the slider button.

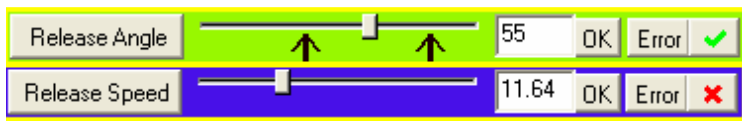


Figure 3: When the error button is pressed, two marks appear either side of the slider button. These marks represent the size of the error. In this case, the release angle contains an error but the release speed is determined. The size and skewness of the error can be changed by moving the position of the two marks.

We see this mechanism as an example of what Papert refers to as a quasi-concrete object (Turkle and Papert, 1991), in the sense that the virtual objects can be manipulated in ways akin to how we learn about material

objects through their use. We are in fact exploiting here an affordance of computational objects to facilitate an intuitive connection to abstract formalisms.

We intend that by playing with the button and markers as illustrated in Figure 3, students will gain an intuitive feel for the role of average (the main slider button), spread (the arrow buttons) and skewness (the degree of symmetry between the two arrows in relation to the main slider button). Students will be able to compare the emergent distributions corresponding to increasing numbers of throws to the settings they have used for their throws. We conjecture students will gain a sense of what aspects of the distribution are directly influenced by their settings and which are not. We hope that they will in this way be able to differentiate between global features of distribution and local randomness.

This approach throws up some interesting questions, which we will hope to be able to address in our analysis when the students use the new design in the next iteration. The students are using controls that determine features (average, spread and skewness) of the emergent distribution but do not entirely define it (the specific results are unpredictable). We wonder whether this is an acceptable resolution of the apparently paradoxical relationship between the determined and the stochastic. It seems that such a resolution is typically only within reach of experts who have constructed probability as a means of operationalising the stochastic. We see echoes of such a resolution in inferential methods, which separate the main effect from random error.

6. References

- 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.
- Davis, B. & Simmt, E. (2003). Understanding Learning Systems: Mathematics Education and Complexity Science, *Journal for Research in Mathematics Education*, 34(2), 137–167.
- Johnson, S. (2001). *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. New York: Simon & Schuster.
- Noss, R. & Hoyles, C. (1996). *Windows on Mathematical Meanings: Learning Cultures and Computers*. Dordrecht, The Netherlands: Kluwer.
- Petrosino, A. J., Lehrer, R., & Schauble, L. (2003). Structuring Error and Experimental Variation as Distribution in the Fourth Grade. *Mathematical Thinking and Learning*, 5(2&3), 131-156.
- 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, May 1998.
- Pratt, D. & Noss, R. (2002): The Micro-Evolution of Mathematical Knowledge: The Case of Randomness, *Journal of the Learning Sciences*, 11(4), 453-488.
- Prodromou, T. (2004). Distribution as Emergent Phenomenon. *Proceedings of the British Society for Research into Learning Mathematics*, Vol. 24(1), 49-54.
- Resnick, M. (1991). Overcoming the Centralised Mindset: Towards an Understanding of Emergent Phenomena. In I. Harel & S. Papert (Eds.), *Constructionism* (pp. 205–214). Norwood, NJ: Ablex.
- Turkle, S. & Papert, S. (1991). Epistemological Pluralism and the Revaluation of the Concrete. In I. Harel & S. Papert (Eds.), *Constructionism* (pp. 161-192). Norwood, New Jersey: Ablex.
- Wilensky, U. (1997). What is Normal Anyway? Therapy for Epistemological Anxiety. *Educational Studies in Mathematics. Special Edition on Computational Environments in Mathematics Education*. Noss R. (Ed.), 33(2), 171-202.
- Wilensky, U. (1999). *NetLogo*. Evanston, IL. Centre for Connected Learning and Computer-Based Learning, Northwestern University.